

Univerzita Karlova v Praze

Přírodovědecká fakulta

Studijní program: Biologie

Studijní obor: Genetika, molekulární biologie a virologie



Bc. Karolína Filipi

Mitogenomická fylogeografie a adaptivní evoluce norníka rudého (*Clethrionomys glareolus*)

Mitogenomic phylogeography and adaptive evolution of the bank vole (*Clethrionomys glareolus*)

Diplomová práce

Vedoucí závěrečné práce: RNDr. Petr Kotlík, Ph.D. (ÚŽFG AV ČR, v. v. i.)

Praha, 2015

Prohlášení:

Prohlašuji, že jsem závěrečnou práci zpracovala samostatně a že jsem uvedla všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

V Praze, 10. 4. 2015

Karolína Filipi (roz. Bendová)

Poděkování

Mé největší díky náleží především členům Laboratoře molekulární ekologie na ÚŽFG AV ČR, v. v. i., v Liběchově. Jmenovitě pak mému školiteli RNDr. Petru Kotlíkovi, Ph.D., za ochotné konzultace a zájem o mou práci, ale především za obrovskou podporu a „pošťouchnutí“ do psaní článků, žádosti o grant a účasti na konferenci, na které bych si sama nikdy netroufla, RNDr. Silvii Markové, Ph.D., za cenné rady a asistenci při práci v laboratoři a boji s počítačovou technikou stejně jako za neuvěřitelnou podporu a zájem o mou práci, Bc. Janě Kopecké za pomoc nejen při zpracovávání vzorků z klonování, Bc. Michaele Strážnické za její nakažlivé nadšení a dodání energie, když mně samotné scházela, a Prof. Jeremymu Searle (Cornell University), ve spolupráci s nímž jsme zpracovávali výsledky mé práce do podoby článků.

Nemalý podíl na dokončení této práce měli také mí rodiče. Nejsem si jistá, nakolik jsem jim vyjadřovala vděk v průběhu svého (poněkud delšího, než jsem plánovala) studia, tedy využiju prostoru zde: Díky, mami, díky, tati. Vaše pochopení a podpora ve studiu pro mě moc znamenaly a znamenají.

Rovněž chci poděkovat svému muži za ohromnou toleranci, vstřícnost, pochopení a péči v době, kdy jsem si již tak dost náročné uzavírání ročníku ještě více zkomplikovala, a v neposlední řadě také za ochotu být mým největším laickým kritikem a pokusným králíčkem, když došlo na přípravy prezentací pro odbornou veřejnost. I když jsem pokaždé myslela, že vyletím z kůže, když jsi protáhl obličej a řekl „Promiň, ale co *tím vlastně chceš říct?*“, byly to právě Tvé poznámky, které mi pomohly udělat prezentace tak, aby byly srozumitelné. Díky, Járo.

Nemohu opomenout ani díky za finanční podporu našeho projektu prostřednictvím grantů GAČR (P506-11-1872) a IGA 2013 (ÚŽFG/13/10), díky níž jsme mohli realizovat takto rozsáhlá sekvenování velkého množství vzorků, a institucionální podporu RVO 67985904 ÚŽFG AV ČR, v. v. i.

Abstrakt

Práce je součástí projektu sekvenování genomu a transkriptomu norníka rudého (*Clethrionomys glareolus*). Význam selekce v evoluci mitochondriální DNA (mtDNA) je předmětem časté diskuse; zatímco některé studie nenašly důkazy, že by selekce významně ovlivnila fylogeografii studovaných druhů, jiné studie naopak adaptivní evoluci mtDNA přikládají velký význam. Norník rudý je hlavním modelem, na kterém studujeme adaptaci na klimatické změny. Jako v případě jiných druhů byla dosud fylogeografie norníka hodnocena na základě variability malé části mtDNA. Cílem mé práce ale bylo osekvenovat celou kódující část mitochondriálního genomu zástupců hlavních linií mtDNA norníka s využitím Sangerovy metody a technologie Illumina, a na příkladu tohoto druhu zhodnotit význam selekce a adaptace v evoluci a fylogeografii. Ačkoli adaptivní evoluce v mtDNA patrně nehrála klíčovou roli při postglaciální kolonizaci Evropy, známky adaptace v mtDNA norníka nalézt můžeme – identifikovala jsem záměnu aminokyseliny s možným funkčním významem a některé populace nacházející se na okrajích areálu rozšíření (severním a jižním) vykazují signifikantní změny fyzikálně-chemických vlastností proteinů kódovaných v mtDNA.

Klíčová slova:

adaptace, molekulární evoluce, populační genetika, sekvence DNA, mitochondriální DNA, mtDNA, cDNA, transkriptom, sekvenování, next generation sequencing, Illumina, savci, obratlovci, Myodes, geografie

Abstract

This thesis is a part of the project aimed at sequencing the genome and transcriptome of the bank vole (*Clethrionomys glareolus*). The role of natural selection in the evolution of mitochondrial DNA (mtDNA) has been subject to much discussion; while some studies did not provide evidence that selection affected the phylogeography of the studied species, other considered adaptive evolution important. The bank vole is the key model we use to study the adaptation to climate change. As with other species, the phylogeography of the bank vole has been based on the variation of a small part of mtDNA. The goal of the thesis was to sequence the entire mitochondrial genome for representatives of all main mtDNA lineages of the bank vole using the Sanger and Illumina technologies, and to assess the role of selection and adaptation in the evolution and phylogeography of this species. The adaptive evolution in mtDNA probably was not the main driving force during the postglacial colonization of Europe. However, signatures of adaptive evolution have been found – an amino acid change with possible functional consequences in one gene and an excess of radical changes in physical-chemical properties of amino acids in populations at the latitudinal (northern and southern) extremes of the bank vole distribution.

Key words:

adaptation, molecular evolution, population genetics, DNA sequence, mitochondrial DNA, mtDNA, cDNA, transcriptome, next generation sequencing, Illumina, mammals, animals, vertebrates, Myodes, geography

Obsah

| | |
|--|----|
| Poděkování | 4 |
| Abstrakt a klíčová slova | 5 |
| Abstract and Key Words | 6 |
| Seznam zkratk | 9 |
| 1. Úvod | 13 |
| 2. Cíle práce | 15 |
| 3. Přehled literatury | 16 |
| 3.1. Molekulární ekologie, fylogenetika a fylogeografie | 16 |
| 3.2. Využití mtDNA jako molekulárního markeru | 17 |
| 3.2.1. Selektivní neutralita mtDNA | 18 |
| 3.2.2. Jaderné mitochondriální pseudogeny (<i>numts</i>) | 20 |
| 3.3. Mitochondrie | 22 |
| 3.3.1. Objev a původ názvu | 22 |
| 3.3.2. Struktura mitochondrií | 22 |
| 3.3.3. Mitochondriální genom | 23 |
| 3.3.4. Replikace mtDNA | 24 |
| 3.3.5. Funkce mitochondrií v buňce | 25 |
| 3.3.6. Původ mitochondrií | 30 |
| 3.4. Modelový organismus – norník rudý | 31 |
| (<i>Clethrionomys glareolus</i>) | 31 |
| 3.4.1. Fylogeografie norníka | 32 |
| 3.4.1.1. Mitochondriální linie norníka v Evropě | 33 |
| 3.4.1.2. Norník jako klíčový model postglaciální kolonizace Evropy temperátními druhy | 34 |
| 3.4.1.3. Linie mtDNA norníků v Británii | 36 |
| 4. Materiál a metody | 39 |
| 4.1. Výběr jedinců ke studii | 39 |
| 4.2. Izolace DNA | 39 |
| 4.3. Amplifikace mtDNA – long-range PCR | 39 |
| 4.3.1. Primery | 40 |
| 4.3.2. Použité reakční směsi a polymerázy | 41 |
| 4.3.3. PCR programy | 42 |
| 4.4. Gelová elektroforéza | 42 |
| 4.5. Purifikace PCR produktů | 43 |
| 4.6. Primer walking – první referenční sekvence | 43 |
| 4.7. Sekvenování metodou Sangera | 44 |

| | | |
|-----------|---|----|
| 4.8. | RNA-Seq norníků z Velké Británie | 45 |
| 4.9. | Alignment mitogenomů | 46 |
| 4.10. | Amplifikace a klonování sekvencí <i>cob</i> | 46 |
| 4.10.1. | Primery | 47 |
| 4.10.2. | DNA polymeráza a master mix | 47 |
| 4.10.3. | PCR program | 47 |
| 4.10.4. | Purifikace a sekvenace | 48 |
| 4.10.5. | Klonování a sekvenace inzertu | 48 |
| 4.11. | Fylogenetické analýzy | 50 |
| 4.11.1. | Testy neutrální evoluce | 50 |
| 4.11.1.1. | TreeSAAP | 50 |
| 4.11.1.2. | GA-Branch | 50 |
| 4.11.1.3. | SLAC/REL/FEL | 51 |
| 4.11.1.4. | PRIME | 51 |
| 4.11.2. | Konstrukce fylogenetických stromů | 52 |
| 4.11.2.2. | PartitionFinder – hledání nejlepšího schématu evoluce sekvencí | 52 |
| 4.11.3. | Analýzy ML se zařazením výsledků z partic | 53 |
| 4.11.4. | Bootstrap analýzy | 54 |
| 4.11.5. | Bayesiánské fylogenetické analýzy - MrBayes | 55 |
| 5. | Výsledky | 56 |
| 5.1. | Amplifikace mtDNA pomocí long-range PCR | 56 |
| 5.2. | Sekvence mitogenomů | 60 |
| 5.2.1. | Vytvoření referenční sekvence norníka rudého (primer walking) | 60 |
| 5.2.2. | Získání kompletních mitogenomů všech mtDNA linií (Sangerovo sekvenování) | 61 |
| 5.3. | Fylogenetické analýzy | 62 |
| 5.4. | Analýza sekvencí <i>cob</i> | 65 |
| 5.5. | Testy neutrální evoluce | 67 |
| 6. | Diskuse | 69 |
| 6.1. | Technická stránka práce | 69 |
| 6.2. | Mitogenomická fylogeografie norníka rudého v Evropě | 70 |
| 6.3. | Baskická linie – pouze artefakt? | 73 |
| 6.4. | Mitochondriální linie norníků v Británii | 76 |
| 6.5. | Testy adaptivní evoluce | 78 |
| 7. | Shrnutí | 80 |
| 8. | Seznam použité literatury | 81 |
| 9. | Přílohy Chyba! Záložka není definována. | |

Seznam zkratek

Při popisu mitochondriálních genů využívám zkratky užívané v programu MITOS (Bernt et al. 2013), který vychází z nomenklatury navržené Jeffreym Boore (Boore 2006), jednak z důvodu konzistence s nomenklaturou používanou v mých člancích (Bendová et al. 2014, Filipi et al. 2014), a také proto, že tento systém označení odkazuje na endosymbiotický původ mitochondrií a Boore ho navrhuje jako univerzální a jednoznačný anotační systém (což je věc, která dosud při popisu organelárních genomů chyběla).

| | |
|------------------|---|
| 12S rRNA | viz rrnS |
| 16S rRNA | viz rrnL |
| acetylCoA | acetyl koenzym A |
| AIC | Akaikeho informační kritérium (angl. <i>Akaike Information Criterion</i>) |
| AICc | upravené Akaikeho informační kritérium (angl. <i>Akaike Information Criterion Corected</i>) |
| AIF | faktor indukující apoptózu (angl. <i>apoptose inducing factor</i>) |
| Ala | alanin |
| AMK | aminokyselina |
| atp6 | podjednotka 6 ATPázové pumpy |
| atp8 | podjednotka 8 ATPázové pumpy |
| BAK | pro-apoptický protein z rodiny BCL-2 (angl. <i>Bcl-2 homologous antagonist/killer</i>) |
| BAX | pro-apoptický protein z rodiny BCL-2, antagonist BCL-2 (angl. <i>Bcl-2-associated X protein</i>) |
| BCL-2 | rodina pro a anti-apoptických proteinů/ anti-apoptický protein z rodiny BCL-2 (angl. <i>B-cell lymphoma 2</i>), antagonist Bax |
| BIC | Bayesiánské informační kritérium (angl. <i>Bayesian Information Criterion</i>) |
| cDNA | vlákno DNA získané reverzní transkripcí z mRNA (angl. <i>copy DNA</i>) |
| cob | cytochrom b |
| cox1 | cytochrom oxidázová podjednotka 1 |
| cox2 | cytochrom oxidázová podjednotka 2 |
| cox3 | cytochrom oxidázová podjednotka 3 |
| D-Loop | nekódující oblast v mtDNA, která je při replikaci vytlačována nově syntetizovaným vláknem (angl. <i>Displacement Loop</i>) |
| dN/dS | někdy také označováno ω ; poměr nesynonymních a synonymních záměn v kódující DNA |

| | |
|--------------------------------|---|
| EndoG | endonukleáza G |
| FEL | metoda ML k odhadu ω (dN/dS); počítá ω zvlášť pro každou pozici v alignmentu (z angl. <i>Fixed Effects Likelihood</i>) |
| FMN | oxidovaný flavinmononukleotid |
| GTR+G | model evoluce DNA (angl. <i>Generalised Time Reversible + gamma correction</i>) |
| GTR+G+I | model evoluce DNA (angl. <i>Generalised Time Reversible + gamma correction + invariant sites</i>) |
| HKY+I | model evoluce DNA (angl. <i>Hasegawa, Kishino a Yano + invariant sites</i>) |
| LGM | vrchol poslední doby ledové (angl. <i>Last Glacial Maximum</i>) |
| MCMC | algoritmus Markov Chain Monte Carlo k odhadu aposteriorní pravděpodobnosti využívaný v Bayesiánských metodách |
| MCMCMC (MC³) | algoritmus Metropolis Coupled Markov Chain Monte Carlo k dosažení lepších výsledků MCMC |
| ML | statistická metoda, která vybírá nejvěrohodnější hypotézu vzhledem k datům (angl. <i>Maximum Likelihood</i>) |
| mtDNA | mitochondriální DNA |
| MTMP | mitochondriální pór (angl. <i>mitochondrial permeability transition pore</i>) |
| nad1 | NADH-dehydrogenázový komplex, podjednotka 1 |
| nad2 | NADH-dehydrogenázový komplex, podjednotka 2 |
| nad3 | NADH-dehydrogenázový komplex, podjednotka 3 |
| nad4 | NADH-dehydrogenázový komplex, podjednotka 4 |
| nad4l | NADH-dehydrogenázový komplex, podjednotka 4L |
| nad5 | NADH-dehydrogenázový komplex, podjednotka 5 |
| nad6 | NADH-dehydrogenázový komplex, podjednotka 6 |
| NGS | sekvenování nové generace (angl. <i>Next Generation Sequencing</i>) |
| NNI | metoda k optimalizaci vnitřní topologie fylogenetických stromů (angl. <i>Nearest Neighbour Interchanges</i>) |
| Numts | jaderné pseudogeny původem z mitochondriální DNA (angl. <i>Nuclear Mitochondrial Pseudeogenes</i>) |
| o_H | počátek replikace těžkého vlákna mtDNA (angl. <i>origin of heavy replication</i>) |
| o_L | počátek replikace lehkého vlákna mtDNA (angl. <i>origin of light replication</i>) |
| ori | počátek replikace, většinou používáno u kruhových molekul DNA (angl. <i>origin of replication</i>) |

| | |
|----------------------|--|
| <i>polyA</i> | polyadenylovaný 3' konec mRNA |
| <i>konec</i> | metody k určení významnosti AMK záměny na základě změny fyzikálně-chemických vlastností v místě záměny (angl. <i>PRoperty Informed Models of Evolution</i>) |
| <i>PRIME</i> | |
| <i>RAxML</i> | program pro rychlou postupnou rekonstrukci fylogenetických stromů (angl. <i>Randomized Axelerated Maximum Likelihood</i>) |
| <i>REL</i> | metoda ML k odhadu ω (dN/dS) uvažující náhodné rozložení záměn podél fylogenie (z angl. <i>Random Effects Likelihood</i>) |
| <i>RITOLS</i> | model replikace mtDNA (angl. <i>RNA Incorporated ThroughOut the Lagging Strand</i>) |
| <i>ROS</i> | reaktivní metabolity kyslíku (angl. <i>Reactive Oxygen Species</i>) |
| <i>rrnL</i> | 16 s rRNA, velká podjednotka mitochondriálního ribosomu |
| <i>rrnS</i> | 12 s rRNA, malá podjednotka mitochondriálního ribosomu |
| <i>SBS</i> | metoda sekvenování (angl. <i>Sequencing By Synthesis</i>) |
| <i>SDM</i> | model replikace mtDNA (angl. <i>Strand Displacement Model</i>) |
| <i>SLAC</i> | metoda ML k odhadu ω (dN/dS) pracující se skutečným počtem všech záměn vyskytujících se podél fylogenie (angl. <i>Single Likelihood Ancestor Counting</i>) |
| <i>SPR</i> | metoda optimalizace při hledání nejlepšího fylogenetického stromu (angl. <i>Subtree Pruning and Regrafting</i>) |
| <i>Thr</i> | threonin |
| <i>trnA</i> | tRNA - alanin |
| <i>trnC</i> | tRNA - cystein |
| <i>trnD</i> | tRNA - kyselina asparágová |
| <i>trnE</i> | tRNA - glutamová kyselina |
| <i>trnF</i> | tRNA - fenylalanin |
| <i>trnG</i> | tRNA - glycin |
| <i>trnH</i> | tRNA - histidin |
| <i>trnI</i> | tRNA - izoleucin |
| <i>trnK</i> | tRNA - lyzin |
| <i>trnL1</i> | tRNA - leucin (antikodon tag) |
| <i>trnL2</i> | tRNA - leucin (antikodon taa) |
| <i>trnM</i> | tRNA - methionin |

| | |
|----------------------------|------------------------------|
| <i>trnN</i> | tRNA - asparagin |
| <i>trnP</i> | tRNA - prolin |
| <i>trnQ</i> | tRNA - glutamin |
| <i>trnR</i> | tRNA - arginin |
| <i>trnS1</i> | tRNA - serin (antikodon gtc) |
| <i>trnS2</i> | tRNA - serin (antikodon tga) |
| <i>trnT</i> | tRNA - threonin |
| <i>trnV</i> | tRNA - valin |
| <i>trnW</i> | tRNA - tryptofan |
| <i>trnY</i> | tRNA - tyrozin |
| ω | viz dN/dS |

1. Úvod

Od konce osmdesátých let je mitochondriální DNA (mtDNA), respektive vybrané geny v ní, hojně využívána ve fylogenetických a fylogeografických studiích (nejen) savců, a je tomu tak oprávněně. (Pozn.: Pro zjednodušení dále v textu uvádím pouze „mtDNA“, přičemž pokud není řečeno jinak, uvedené informace se týkají savčí mtDNA.) Jak velice pěkně shrnují Avise et al. (1987) ve svém klíčovém článku, mtDNA totiž do značné míry splňuje většinu požadavků na ideální fylogenetický marker.

První devizou mtDNA je její kompaktní velikost (cca 15-17 kbp) a přítomnost v buňce v mnoha kopiích (především u buněk silně metabolicky aktivních se jedná až o řády tisíců kopií na buňku). Díky tomu je snadné molekulu mtDNA izolovat či amplifikovat vybrané geny i z malého množství zdrojového materiálu. Další výhodou je, že mtDNA má oproti jaderné DNA vysokou mutační rychlost a především téměř nepodléhá rekombinaci. Tyto dvě její vlastnosti jsou nedocenitelné při fylogenetických studiích, neboť umožňují velmi přesné stanovení příbuzenských vztahů jak mezidruhově, tak vnitrodruhově. Nelze také opomenout význam dědičnosti mtDNA po maternální linii - při zasazení do geografického kontextu tato umožňuje sledovat původ mitochondriálních linií, čehož se využívá při fylogeografických studiích. Z nejznámějších studií mtDNA vzpomeňme například objasňování otazníků okolo geografického původu člověka a hledání „mitochondriální Evy“ (viz Cann et al. 1987).

O mtDNA se dlouhou dobu mluvilo jako o selekčně neutrálním fylogenetickém markeru, respektive se předpokládalo, že geny v mtDNA budou spíše pod vlivem negativní selekce, než že by mohly být předmětem adaptivní evoluce. Tato domněnka není úplně neoprávněná – dle teorie neutrální evoluce (Kimura, 1983) většina mutací v DNA nemá funkční význam.

Fakt, že mtDNA obsahuje 13 genů kódujících protein (PCG), 2 kódující rRNA a 22 kódujících tRNA, které se významně podílí na metabolismu buňky, a záměna v nich se může velice významně podepsat na funkci výsledného proteinu či rRNA, je sice spíš argumentem pro teorii mtDNA jako neutrálního markeru (nevýhodné mutace jsou velice rychle odstraněny selekcí). Na druhou stranu však není důvod se domnívat, že krom mutací neutrálních či s negativním efektem nemůže v mtDNA dojít i k záměnám, jež mohou mít pozitivní vliv na zdatnost (fitness) jedince.

Zároveň je třeba podotknout, že velká většina studií, které za posledních několik desítek let s mtDNA jako markerem pracují, využívá pouze malou, předem vybranou část mtDNA. Tento výběr se liší v závislosti na druhu, na kterém je studie prováděna; velmi oblíbený je například gen pro cytochrom b (*cob*), hojně využívána je také nekódující hypervariabilní oblast (tzv. D-Loop). Ačkoli se tyto markerové oblasti mtDNA v minulosti osvědčily, celý mitochondriální genom obsahuje výrazně větší množství genů a funkčních RNA, a je také jako celek mnohem informativnější. Rozvoj nových sekvenačních technik (tzv. Next Generation Sequencing, zkráceně NGS) nám dnes navíc umožňuje velmi snadné získání rozsáhlých úseků DNA, navíc mtDNA je pro tyto účely extrémně vhodná, protože neobsahuje introny, což celou analýzu velice zjednodušuje (např. sekvenujeme-li transkriptom, viz dále).

Rozhodla jsem se tedy provést tuto fylogenetickou studii, zaměřenou na roli adaptivní evoluce mtDNA v postglaciální kolonizaci Evropy po poslední době ledové, na celkové mtDNA. Jako modelový organismus jsem zvolila norníka rudého (*Clethrionomys glareolus*), hlodavce z čeledi hrabošovitých, který je typickým zástupcem druhů žijících v lesích mírného pásma. Norník rudý je jako modelový druh ve fylogenetických studiích zabývajících se postglaciální kolonizací Evropy velmi populární (Deffontaine et al. 2005; Kotlík et al. 2006; Searle et al. 2009). Důvodem je především široký areál rozšíření, který sahá od jihu Evropy až za polární kruh - v celé Evropě bylo na základě dat z *cob* popsáno 8 mitochondriálních linií¹ norníků s různou geografickou distribucí. Dosavadní studie však předpokládaly, že evoluce mtDNA norníka byla neutrální. Norník je ale zajímavý i jako učebnicový příklad mezidruhově introgrese mtDNA (tzv. mtDNA capture) – tato byla popsána u skandinávských populací, které nesou mtDNA příbuzného norníka tajgového *C. rutilus* (Tegelström 1987), přičemž není vyloučeno, že tato introgrese měla význam v adaptaci druhu (Boratyński et al. 2011).

Mým cílem tedy bylo jednak zjistit, zda linie popsané na základě jednoho genu (*cob*) (Deffontaine et al. 2005, Kotlík et al. 2006, Colagnello et al. 2012) budou odpovídat liniím popsaným na základě kompletního mitogenomu, a rovněž otestovat, zda některé z těchto linií nevykazují v mtDNA známky adaptivní evoluce. A protože mitogenom norníka rudého dosud nebyl kompletně osekvenován a popsán – nejbližším příbuzným popsaným druhem byl korejský norník *C. regulus* (Kim a Park 2012) – musela jsem pro účely studie nejprve osekvenovat celý mitogenom norníka rudého *de novo*.

¹ Někdy se jako synonymum pro linii mtDNA používá také výraz „klád“ (z anglického *clade*). Nejčastěji se s tímto termínem setkáme při popisu fylogenetických stromů.

2. Cíle práce

1. Osekvenovat *de novo* metodou primer walking kompletní mitogenom norníka rudého.
2. Získanou sekvenci využít k návrhu sekvenačních primerů použitelných pro další jedince.
3. Dalším sekvenováním (metodami Sangera a Illumina) získat kompletní sekvenci mitogenomu pro zástupce všech dosud popsanych linií v Evropě.
4. Ze získaných sekvencí zkonstruovat fylogenetické stromy a zjistit, zda data z kompletní mtDNA potvrdí linie popsané na základě *cob*.
5. Mitochondriální linie, identifikované na základě fylogeneze kompletních mitogenomů, otestovat na známky adaptivní evoluce.
6. Odpovědět na otázku, zda je mtDNA norníka předmětem adaptivní evoluce (a pokud ano, ve kterých genech).

3. Přehled literatury

3.1. Molekulární ekologie, fylogenetika a fylogeografie

Dříve, než se začnu podrobněji zabývat fylogenetikou a fylogeografií, dovolím si krátký úvod do molekulární ekologie.

Molekulární ekologie se, stejně jako „tradiční“ ekologie, zabývá vztahy mezi organismy a jejich prostředím, je však specifická v tom, že k zodpovězení svých otázek využívá molekulární metody, které analyzují biologické makromolekuly: DNA, RNA a proteiny.

V molekulární ekologii je využíváno několik hlavních metodických přístupů, které se však vzájemně nevylučují, ale spíše doplňují. Pro zasazení studovaného problému do širšího kontextu je dokonce velmi žádoucí je kombinovat.

Prvním možným přístupem v molekulární ekologii je studium konkrétního genu či proteinu, podrobné zkoumání variability mezi jedinci a zjišťování, zda a jaký má tato variabilita význam pro organismus, například v embryogenezi, metabolismu, fyziologii atd. (např. Kotlík et al. 2014).

Další směr, fylogenetika, pak využívá genetické variability pouze jako markeru, a zkoumá evoluční příbuznost mezi skupinami organismů (druhů či populací) pomocí molekulárních metod (především srovnáváním podobnosti na úrovni DNA či proteinu). Výhodou použití molekulárních metod ve fylogenetice je, že dovedou odlišit homologii od analogie, což může být zásadní problém, pokud používáme pouze morfologické znaky. Samotný pojem fylogenetika je odvozen od dvou řeckých slov – podstatného jména *fylé* (φυλή) či *fylon* (φύλον), jež se dá přeložit jako „kmen“, „druh“ či „klan“, a přídavného jména *genetikós* (γενετικός), jež je odvozeninou ze slova *genesis* (γένεσις), znamenajícím „původ“, „počátek“, „zdroj“ či „narození“. Výsledkem fylogenetických analýz je fylogeneze, neboli hypotéza o evoluční historii a vzájemné příbuznosti zkoumaných jednotek (alel genu, taxonů apod.). Nejběžnějším a graficky nejsrozumitelnějším zobrazením fylogeneze je tzv. fylogenetický strom, jehož tvar se označuje jako topologie a je dán uspořádáním uzlů (anglicky *nodes*) a větví (*branches*). Ilustruje evoluční historii (čili pořadí, v jakém se jednotlivé taxony oddělovaly od společného předka), kde délka větví odráží genetickou vzdálenost, tj. míru realizované genetické změny – evoluce.

Fylogeografie, kterou se zabývá má práce, je v podstatě kombinací obou výše zmíněných přístupů a zároveň jejich zasazením do geografického kontextu. Zkoumá genetické vztahy mezi populacemi v kontextu jejich geografické distribuce, na základě čehož hodnotí historii populací a jejich rozšíření, a nověji také studuje funkční vliv variability v genech.

3.2. Využití mtDNA jako molekulárního markeru

Jak již bylo zmíněno výše, mtDNA je oprávněně jedním z nejčastěji užívaných markerů ve fylogenetických a fylogeografických studiích. Největší rozmach užívání mtDNA nastal po otištění článku Avise et al. (1987), v němž autoři formálně uvádějí disciplínu fylogeografie a definují požadavky na ideální molekulární systém. Takový systém by měl:

- 1) být jasně definovaný a přítomný u široké škály organismů, aby umožňoval homologní srovnání mezi různými druhy,
- 2) být snadno izolovatelný a použitelný pro analýzy,
- 3) mít jednoduchou genetickou strukturu bez repetitivní DNA, mobilních genetických elementů, pseudogenů a intronů,
- 4) být předáván do další generace bez účasti rekombinace či jiných genetických přestaveb,
- 5) nést vícero kvalitativních znaků, jejichž evoluce může být stanovena pomocí parsimonních kritérií,
- 6) pro účely mikroevolučních (vnitrodruhových) analýz mít takovou mutační rychlost, která způsobí, že v něm běžně vznikají nové mutace.

MtDNA tyto požadavky do překvapivě velké míry naplňuje. Její malá velikost, vyšší stabilita (ve srovnání s jadernou DNA) a přítomnost v mnoha kopiích v buňce umožňuje snadnou izolaci. Genetická struktura mtDNA s velmi malým podílem nekódujících oblastí a absencí intronů je u živočichů silně konzervativní (Gissi et al. 2008). Zároveň se v ní střídají konzervativní úseky DNA s úseky variabilnějšími, což je ideální pro návrh PCR primerů.

Oproti jaderné DNA má mtDNA vysokou mutační rychlost (Brown et al. 1979, Vawter a Brown, 1986), jež má za následek její vysokou vnitro- i mezidruhovou variabilitu, a tedy použitelnost nejen pro analýzy nejen velmi vzdálených druhů (např. *cox1* je dnes standardně využíván jako tzv. barcoding gene, tedy gen, pomocí něhož je možné vzájemně odlišit různé druhy), ale rovněž pro analýzu jedinců v rámci jednoho druhu (na základě podobnosti v mtDNA lze identifikovat různé maternální linie).

Vyšší frekvence mutací v mtDNA byla dříve přikládána velké koncentraci ROS (*reactive oxygen species*, reaktivních metabolitů kyslíku) v mitochondriích a méně výkonným opravným mechanismům při replikaci (Driggers et al. 1993). Song et al. (2005) však ukazují na jiný možný mechanismus, způsobující vyšší chybovost jinak velmi přesné DNA polymerázy γ , a to nerovnováhu v koncentraci dNTP v mitochondriích.

Široce rozšířený a mezi vědeckou komunitou poměrně dlouho uznávaný byl také předpoklad, že mtDNA je selekčně neutrální nebo pod mírnou negativní selekcí (Ohta, 2002).

Vzhledem k tomu, že mtDNA se dědí po maternální linii, tedy klonálně, a při přenosu do další generace se navíc uplatňuje silný bottleneck (Bergstrom a Pritchard, 1998, Marchington et al. 1997) a v jeho důsledku i genetický drift (dochází k přechodnému snížení počtu mitochondrií), většina organismů je, co se týká mtDNA, homoplasmická (má ve svých buňkách pouze jeden haplotyp mtDNA). Dlouhou dobu se také předpokládalo, že mtDNA není ovlivněna rekombinací - bylo popsáno mnoho mechanismů eliminace paternální mtDNA u různých organismů (Moses 1961, Sutovsky 1999, Ursprung a Schabtach 1965). Právě absence rekombinace a klonální dědičnost mtDNA je pro účely fylogenetických analýz velice výhodná, neboť všechna místa na jedné mtDNA mají stejnou genealogickou historii.

Ačkoli se mtDNA po předchozím výčtu může jevit jako dokonalý marker pro studium evoluční historie druhů, nelze k ní takto jednoznačně přistupovat. Stále se totiž objevují další a další práce, které zpochybňují předchozí závěry ohledně jejích vlastností, ať už mluvíme o homoplasmii (zdá se, že mnoho chorob způsobených mutacemi v mitochondriích je spojeno s určitou hladinou heteroplasmie v mtDNA), striktně maternální dědičnosti (u ptáků byly popsány případy přenosu paternální mtDNA do potomstva, viz např. Kvist et al. 2003), nebo nepřítomnosti rekombinace (u ryb; viz Tatarenko a Avise 2007). Pro fylogenetické a fylogeografické analýzy mtDNA obratlovců pak mají asi největší praktický význam pochybnosti o selekční neutralitě mtDNA a přítomnost jaderných mitochondriálních pseudogenů. Proto jim zde věnuji větší prostor, než ostatním otázkám týkajícím se mtDNA.

3.2.1. Selekční neutralita mtDNA

Obecně se předpokládalo, že mtDNA je neutrální nebo téměř neutrální, pokud jde o selekci. Avise et al. (1987) ve svém významném článku uzavřeli problematiku neutrality mtDNA s tím, že většina rozdílů mezi jednotlivými haplotypy mtDNA segregujícími v populaci nejspíše nemá vůbec žádný dopad na fitness jedince (na základě toho, že většina do té doby objevených

záměn se vyskytovala v nekódujících místech či se jednalo o synonymní mutace). Na druhou stranu, mitochondriální geny vykazují vyšší poměry nesynonymních a synonymních (dN/dS) mutací než jaderné geny (Lynch 1996), přičemž právě nesynonymní mutace mají za následek aminokyselinové (AMK) záměny ve výsledném proteinu. Pokud by šlo o takovou záměnu, která výrazně změní například polaritu nebo náboj daného místa v AMK řetězci a v důsledku pak i jeho schopnost účastnit se vazeb a elektrostatických interakcí s jinými AMK ve stejné nebo i jiné molekule či s ligandy, může mít tato záměna funkční význam. A ačkoli je podle neutralistické školy většina pozorovaných substitucí v DNA neutrálních (proti škodlivým mutacím totiž působí poměrně silná negativní selekce), není důvod se domnívat, že by některé nesynonymní substituce v mtDNA nemohly mít vliv na funkci proteinu.

Je také potřeba podotknout, že nejen protein-kódující geny jsou potenciálním terčem selekce – významný vliv na funkci mitochondrií může mít i modifikace translačního systému, tedy mutace v rRNA či tRNA (mutace v tRNA mohou mít za následek např. změnu frekvence využívání určitých kodonů).

Během posledních let se objevují nové články, které poukazují na to, že mtDNA by mohla být předmětem adaptivní evoluce u mnoha různých druhů (Nachman et al. 1996, McClellan et al. 2005, Foote et al. 2011, Soares et al. 2013, Parmakelis et al. 2013). Možnost, že mtDNA může být předmětem selekce, je tedy třeba brát vážně, a je třeba s ní počítat při interpretaci výsledků z analýz. Nižší než očekávaná variabilita mtDNA mezi populacemi totiž většinou ukazuje na nedávné oddělení těchto populací, ale selekce v mtDNA může rovněž způsobit pokles vnitrodruhové variability (Rato et al. 2011), pokud je stejný haplotyp zvýhodněn v obou populacích a dochází mezi nimi ke genovému toku. Pokud pracujeme pouze s daty z mtDNA a nemáme srovnání například s daty z jaderné DNA (která by měla vykazovat podobnou neutrální variabilitu jako mtDNA, s přihlédnutím k odlišné mutační rychlosti), můžeme velmi snadno dojít k mylným závěrům ohledně evoluční historie studovaného organismu.

Významná práce, která podporuje myšlenku, že mtDNA je předmětem adaptivní evoluce, je od francouzských autorů Bazina et al. (2006). Zabývá se velikostí populací a vnitrodruhovou variabilitou v mtDNA a jaderné DNA a byla provedena na extrémně velkém datovém souboru (více než 1600 druhů). Jedním z nejzajímavějších zjištění tohoto týmu bylo, že vnitrodruhová diverzita v mtDNA je překvapivě podobná napříč mnoha živočišnými druhy a není průkazně ovlivněna velikostí populace. Autoři tuto skutečnost vysvětlují tím, že mtDNA opakovaně prochází procesem tzv. *selective sweep* (někdy se používá český termín „selekční vymetení“,

ale já dále budu používat originální anglický). Jde o formu genetického draftu (*genetic draft* nebo také *hitchhiking*, česky „genetické svezení se“), tedy situaci, kdy v populaci dojde ke zvýšení frekvence určité alely díky genetické vazbě s výhodnou mutací. Termín *selective sweep* se vztahuje k případům, kdy dojde v důsledku pozitivní selekce ke snížení variability v DNA v bezprostředním okolí výhodné mutace. Tento jev bude u mtDNA ještě výraznější než u jaderných genů, neboť v mtDNA nedochází k rekombinaci (nebo aspoň ne na takové úrovni jako v jaderné DNA), a proto může dojít až k úplné eliminaci variability v celé mtDNA. Nově vzniklé varianty jsou potom všechny odvozeny od takto zafixovaného haplotypu.

Galtier et al. (2009) jako tři hlavní možné příčiny *selective sweep* jmenují selekci na úrovni jedince (některé varianty mtDNA mohou být výhodnější než jiné – například ty, v nichž se vyskytují mutace, které mají za následek vyšší efektivitu dýchacího systému), dále pak „sobecké“ mutace v mtDNA, tedy takové, které zajistí četnější přenos dané varianty mtDNA do potomstva bez ohledu na funkční dopad na fenotyp (tento jev byl popsán např. u *Saccharomyces cerevisiae* – více MacAlpine et al. 2001) a konečně draft v případech, kdy je mtDNA ve vazbě s jiným genetickým elementem, na který selekce působí. K tomu dochází v případě, kdy je mtDNA předávána do další generace spolu s dalšími genetickými elementy, které jsou děděny maternálně, například symbiotickými mikroby rodu *Wolbachia*, kteří se vyskytují u některých druhů členovců (Turelli et al. 1992), nebo chromozomem W u ptáků (Berlin et al. 2007). Pokud bude působit selekce na některý z těchto elementů, může dojít zároveň i k selekci na úrovni mtDNA.

Byly také zaznamenány případy tzv. *mitochondrial capture* (Shaw 2002, Good et al. 2008, Marková et al. 2013, Boratyński et al. 2011). Jde o jev, kdy se mtDNA charakteristická pro jeden druh vyskytuje u jiného druhu (tedy na cizím genetickém pozadí, co se týká jaderných genů) v důsledku mezidruhového křížení, které mezi nimi proběhlo v minulosti (Avise, 2004). V případě, že na mtDNA působí selekce a haplotyp cizího druhu je zvýhodněn oproti vlastnímu, může dojít až k fixaci tohoto cizího haplotypu. Tato fixace pak může mít za následek, že fylogenetické stromy vytvořené na základě dat z mtDNA neodpovídají evoluční historii (druhy se mohou jevit příbuznější, než ve skutečnosti jsou).

3.2.2. Jaderné mitochondriální pseudogeny (*numts*)

Další možná úskalí fylogenetických analýz pomocí mtDNA představují jaderné mitochondriální pseudogeny, tzv. *numts* (*nuclear mitochondrial pseudogenes*). Jde o kopie

mitochondriálních genů v jaderném genomu. *Numts* se mezi druhy liší, co se týče jejich počtu, délky či pozice na chromozomech, ale jejich výskyt je velmi častý napříč celým spektrem druhů (Bensasson et al. 2001). Odpověď na otázku, proč vlastně jaderné mitochondriální pseudogeny vůbec existují, zůstává stále předmětem diskuse (Doolittle 1998, Berg a Kurland 2000). U většiny eukaryot však v průběhu evoluce dochází k postupné redukci mitochondriálního genomu, a přenos esenciálních genů z mitochondriální do jaderné DNA může být způsobem, jak tyto geny zachovat ve funkční podobě (Adams a Palmer 2003). Samotný mechanismus přenosu ještě není zcela objasněn, ale jako hypotéza se nabízí, že části mtDNA jsou do chromozomální DNA integrovány během oprav zlomů v jaderné DNA nehomologní rekombinací (Blanchard a Schmidt, 1995); objevují se studie, které tuto hypotézu podporují (Yu a Gabriel 1999). Je však potřeba mít na paměti, že přenos mtDNA do jaderné DNA se automaticky nepojí se vznikem funkční kopie přenesených genů v jaderné DNA (už jenom proto, že mtDNA má jiný genetický kód a produkty některých genů by tak nebyly funkční kvůli předčasné terminaci translace).

Fakt, že poměrně běžně dochází k přenosu částí mtDNA do jaderné DNA (viz review Bensasson et al. 2001), má pro fylogenetické analýzy zásadní důsledky. Jak již bylo zmíněno, mtDNA má výrazně vyšší mutační rychlost než jaderná DNA. To znamená, že pokud je do jaderného genomu vnesena určitá část z mtDNA, je zde mnohem lépe „evolučně konzervována“, než v samotné mtDNA. Jinými slovy, jaderný mitochondriální pseudogen přenesený z mtDNA do jaderné DNA bude méně poznamenán novými mutacemi (ve srovnání s mtDNA genem, z něhož je odvozen), a bude se v analýzách jevit podobnější ancestrální sekvenci či jako „molekulární fosilie“. Existují také práce, které ukazují, že opakované translokace mtDNA do jaderné DNA mohou být často také následovány tandemovou duplikací (Lopez et al. 1994, Antunes et al. 2007, Triant a DeWoody 2007). Pokud PCR amplifikaci mitochondriálních genů nepředchází izolace mtDNA a jaderné mitochondriální pseudogeny mají sekvence paralogní k oblastem v mtDNA, kam nasedají PCR primery, může dojít k jejich amplifikaci (zvláště v případech, kdy při přenosu do jádra došlo zároveň i k duplikaci sekvencí vnesených pseudogenů).

Jako příklad toho, jak se amplifikace pseudogenů může podepsat na výsledcích analýz, je možné uvést práci Macholána et al. (2012), kteří v ostrovní populaci myši domácí objevili novou mtDNA linii, zcela odlišnou od ostatních poddruhů myši domácí. Poté, co osekvenovali celý mitochondriální genom jednoho jedince z této populace, se však ukázalo, že domnělá linie

odpovídá sekvenci jaderného mitochondriálního pseudogenu. Tento příklad jasně ukazuje, že pokud nejsou pseudogeny rozpoznány, mohou značně pokřivit výsledky fylogenetických analýz.

3.3. Mitochondrie

3.3.1. Objev a původ názvu

Organely, které dnes označujeme jako mitochondrie, byly poprvé popsány v druhé polovině 19. století, kdy došlo jednak k výraznějšímu uplatnění světelné mikroskopie (firma Carl Zeiss začala sériově vyrábět mikroskopy) a hlavně rozvoji technik barvení preparátů. Během poměrně krátké doby vzniklo mnoho termínů, které popisovaly stejnou strukturu (blepharoblasty, chondriokonta, chondriomity, chondrioplasty, plastochondrie, plastosomy, mitogel, bioblasty a mnoho dalších). Tyto termíny ve většině případů reflektovaly strukturu organely – nejčastěji její zrnkovitost (řecké *chondros*, anglické *grain* či německé *Korn*) nebo válcovitý až vláknitý tvar (latinské *fila*, řecké *mitos*). První použití dnes již zavedeného termínu mitochondrie, který je složeninou dvou řeckých slov – *mitos* (vlákno) a *chondros* (zrno), se pak datuje k přelomu 19. a 20. století a jeho autorem je německý mikrobiolog Carl Benda (1898).

3.3.2. Struktura mitochondrií

Mitochondrie jsou objemově nejrozsáhlejším buněčným kompartmentem. Čím více je buňka metabolicky aktivní, tím větší procento objemu mitochondrie zaujímají – u jaterních buněk, kde se vyskytují 1-2 tisíce mitochondrií na buňku, tak představují až pětinu celkového objemu buňky (Nečas et al. 2000). Jedná se o poměrně dynamické struktury (mohou velice snadno měnit tvar i velikost), které mají nejčastěji válcovitý či vláknitý tvar o délce 3-4 μm a průměru 0,5-1 μm . Jsou navázány na mikrotubuly cytoskeletu. Tato vazba slouží jednak k fixaci mitochondrií na určité místo v buňce (v místech vysoké spotřeby ATP), ale rovněž má regulační funkci při jejich přemísťování a dělení.

Mitochondrie sestávají z vnitřní a vnější membrány. Ty organelu rozdělují na dva prostorově oddělené kompartmenty: mezimembránový prostor, nacházející se mezi vnější a vnitřní membránou, a tzv. matrix – prostor ohraničený vnitřní membránou. Obě membrány se svým složením velmi liší. Vnější membrána obsahuje velké množství porinů, které v ní vytváří kanály, a je tedy velmi propustná pro ionty. Oproti tomu vnitřní membrána je propustná jen velmi málo; velký podíl na tom patrně má vysoký obsah fosfolipidu kardiolipinu (Nečas

et al. 2000). Zatímco vnější membrána je poměrně hladká, vnitřní je bohatě zvlněná a vytváří tzv. mitochondriální krysty. Tyto krysty velmi výrazně zvětšují plochu využitelnou pro syntézu ATP. Právě na vnitřní membráně jsou totiž lokalizovány proteiny, které se účastní elektronového řetězce, a komplex ATP syntetáz. Jsou zde rovněž přítomny transportní proteiny, které vyměňují metabolity pro enzymy matrix. V matrix mitochondrií je pak lokalizována mitochondriální DNA (která bývá přítomna v 5-10 kopiích), ribozomy, tRNA a proteosyntetické enzymy a enzymy účastníci se Krebsova cyklu a β -oxidace mastných kyselin.

3.3.3. Mitochondriální genom

Ačkoli mtDNA rostlin může dosahovat značných délek (kolem 200 kbp), u savců se vyskytuje jako malá, kompaktní kruhová molekula dsDNA o velikosti přibližně 15-17 kbp. Poměr GC/AT párů je zde ve prospěch AT (většinou kolem 60%). Kvůli rozdílné distribuci purinových (těžších) a pyrimidinových (lehčích) bází na jednotlivých vláknech mtDNA byly zavedeny termíny lehké (light, L) vlákno a těžké (heavy, H) vlákno. Zatím není známo, že by tato vlastnost měla pro mtDNA nějaký funkční dopad, ale můžeme ji využít v laboratoři při separaci konkrétního vlákna mtDNA. Tyto termíny se také využívají při popisu umístění genů na mtDNA.

Genetická struktura mtDNA je u zvířat velmi konzervativní; obsahuje jen velmi málo duplikací, žádné introny a velmi krátké nekódující mezigenové oblasti (Gissi et al. 2008). Savčí mtDNA obsahuje 13 protein kódujících genů, které se účastní dýchacího řetězce na vnitřní membráně mitochondrií, 22 genů pro tRNA a 2 geny, kódující rRNA podjednotky mitochondriálního ribozomu (12S rRNA a 16S rRNA). Většina proteinových genů je kódována na těžkém vláknu; výjimku tvoří gen pro ND6 a několik genů pro tRNA.

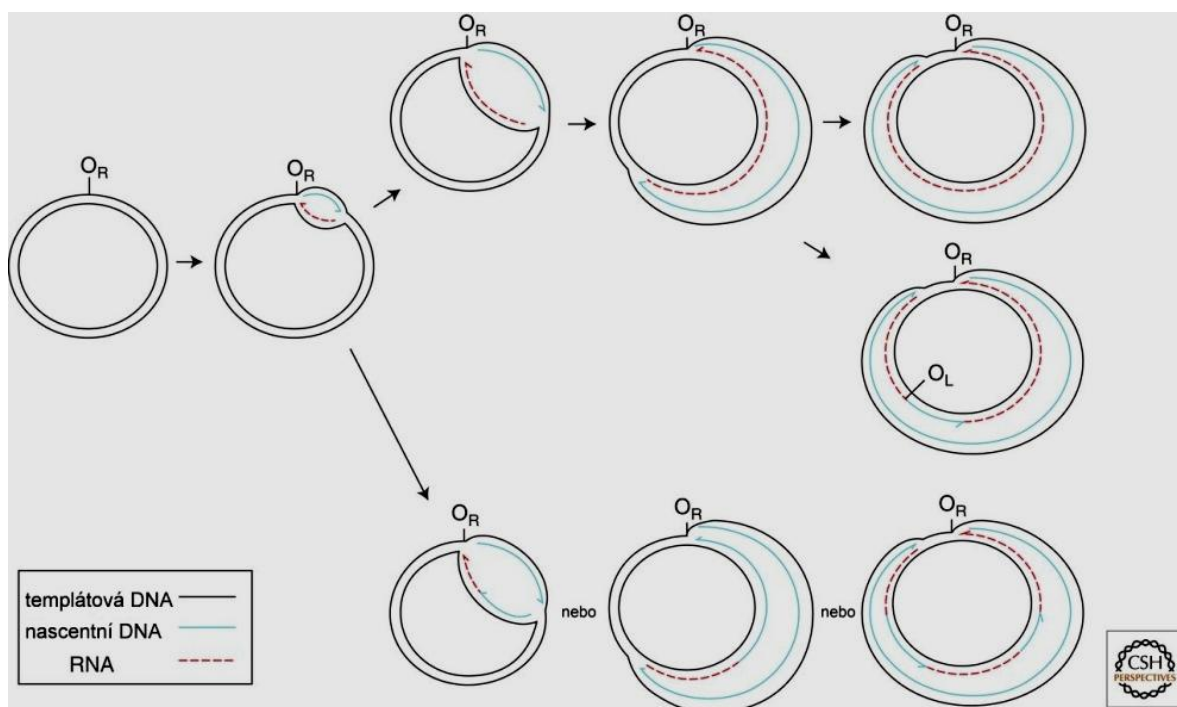
Krom výše zmíněných genů v mtDNA také nalezneme tzv. D-Loop (používá se také český termín *D-smyčka*, *hypervariabilní* nebo také *kontrolní oblast*). Jde o jedinou rozsáhlejší (~1 - 1,5 kbp) nekódující oblast v celém mitogenomu, která je pro svou vysokou variabilitu využívána pro některé typy genetických analýz (využití má jednak ve fylogeografii, ale například také při identifikacích jedinců ve forenzní genetice). Název D-Loop je zkratkou pro *displacement loop* (česky asi nejlépe „vytlačovaná smyčka“), a je odvozen od struktury, kterou tato oblast vytváří při replikaci mtDNA.

3.3.4. Replikace mtDNA

O tom, jak probíhá replikace mtDNA, se stále vedou diskuse; všeobecná shoda zatím panuje jenom ohledně jejího asymetrického průběhu. Donedávna nejvíce uznávaný byl SDM (*Strand Displacement Model*, také známý jako *asynchronní* nebo *asymetrický model replikace*), který navrhli Kasamatsu a Vinograd (1973). Tento model pracuje s tím, že počátky replikace (označované stejně jako u prokaryotních buněk *ori* – zkratka z *origin of replication*) jsou různé pro lehký a těžký řetězec. První se replikuje lehký řetězec. Jeho replikace začíná syntézou RNA primeru právě v oblasti O_H (*ori on the heavy strand* čili počátku replikace na těžkém řetězci), která leží právě v D-smyčce (proto se někdy používá spíše výraz *D-replikace*). S postupující syntézou nového lehkého řetězce je vlákno, na němž leží kontrolní oblast, vytlačováno a vzniká smyčka (odtud *Displacement Loop*). Když syntéza nového těžkého řetězce dosáhne místa, ve kterém leží počátek replikace na lehkém řetězci (O_L), začne se syntetizovat řetězec těžký. Výsledkem jsou pak 2 kruhové molekuly dsDNA, kde jeden řetězec je vždy původní a jeden nový.

Mezi poslední navrhované modely však patří tzv. RITOLS (*RNA Incorporated ThroughOut the Lagging Strand*) model (Holt a Reyes, 2012). Ten vychází z poznatku Yasukawy et al. (2006), kteří potvrdili, že při replikaci mtDNA dochází ke vzniku dvouvláknových RNA/DNA replikačních intermediátů.

RITOLS se v mnohém s SDM shoduje - tvrdí, že iniciace replikace začíná ve dvou počátcích replikace, označovaných O_H pro počátek na těžkém řetězci a O_L pro počátek na lehkém řetězci. Rovněž pracuje s asymetrickým průběhem replikace, kdy jeden (lehký) řetězec DNA je syntetizován dříve (*leading strand*, vedoucí vlákno) a druhý později (tzv. *lagging strand*, opoždující se vlákno), většinou z O_L . Zásadní rozdíl je v tom, že v modelu RITOLS vzniká zároveň s novým lehkým řetězcem také hybridní řetězec RNA/DNA, který zůstává na dosud nereplikovaném templátovém vlákně, a je postupně odstraněn až při syntéze DNA opoždujícího se vlákna (přesný mechanismus replikace tohoto druhého vlákna zatím není objasněn). Pro lepší představu viz obr. 1 (převzato z Holt a Reyes 2012).



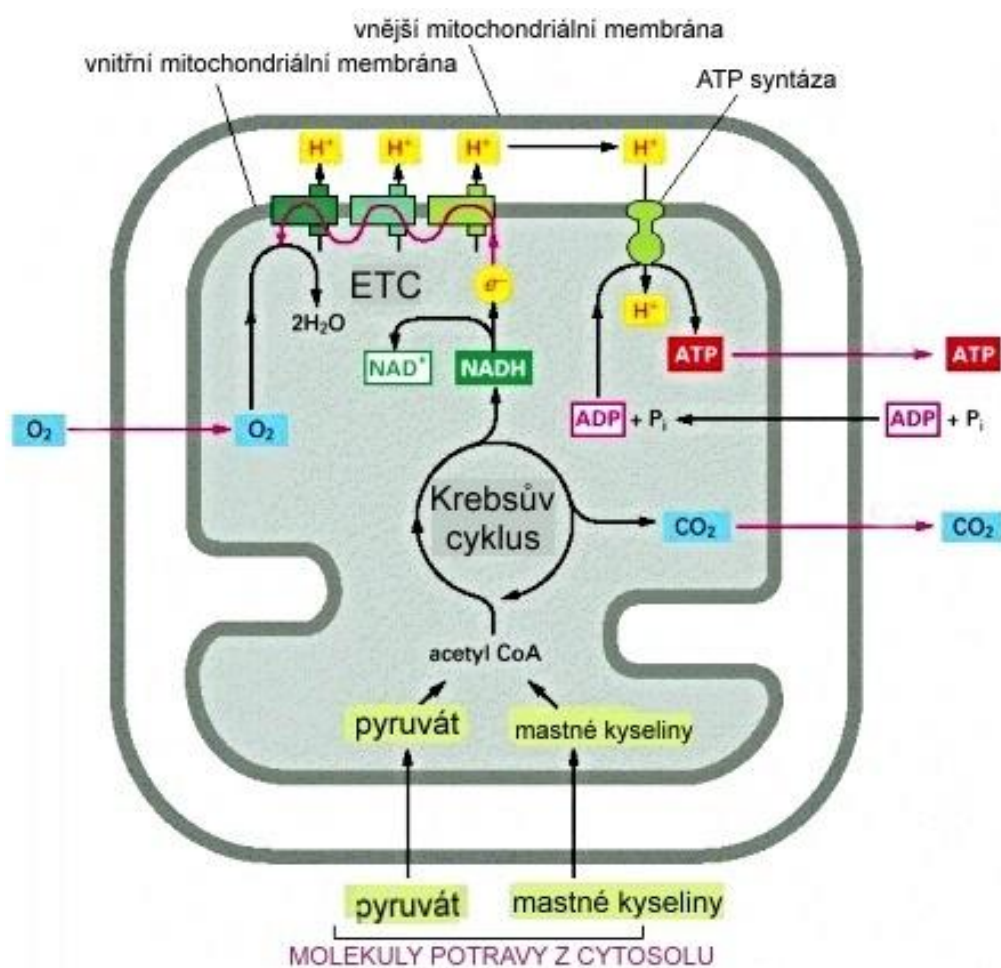
Obr. 1. Diskontinuální model replikace mtDNA s využitím RNA intermediátu – RITOLS $O_R = O_H$. Podle Holt a Reyes 2012.

3.3.5. Funkce mitochondrií v buňce

Funkce mitochondrií v metabolismu eukaryotických buněk byla postupně objasněna až v druhé polovině dvacátého století. S myšlenkou, že mitochondrie (v této práci popisovaných jako *grana*) by se mohly účastnit buněčné respirace, přišel už Warburg (1913), ale ve své době neměl nástroje, aby svou hypotézu experimentálně ověřil. O první izolace mitochondrií pomocí buněčné frakcionace se sice od konce 30. let pokoušel především Claude (1940, 1944, 1946, 1948), který má na vývoji metodiky lví podíl, ale opravdu kvalitní mitochondrie se podařilo získat až Hogeboomovi et al. (1948). Role mitochondrií v buněčném metabolismu tak byla stanovena až později; respirační řetězec popsali Chance a Williams v roce 1956 a chemiosmotickou hypotézu oxidativní fosforylace formuloval Mitchell v roce 1961 (a v roce 1978 za tento objev získal Nobelovu cenu).

Mitochondrie představují největší energetické centrum buňky. V matrix probíhá Krebsův cyklus, na vnitřní membráně potom syntéza ATP pomocí oxidativní fosforylace (neboli toku elektronů z NADH/FADH₂ přes řetězec membránových přenašečů na konečný akceptor, jímž je O₂, a využití elektrochemického potenciálu, který při tomto toku vzniká, k syntéze ATP).

Oba tyto procesy jsou zásadní součástí aerobní respirace, která probíhá celkem ve čtyřech krocích. Jako první dochází ke glykolýze, čili přeměně jedné molekuly glukózy na 2 molekuly pyruvátu (za redukce NAD^+ na NADH). Vzniklý pyruvát je pak dále přeměněn (opět za redukce NAD^+ na NADH) na acetylkoenzym a (acetyl CoA). Ten potom vstupuje do Krebsova cyklu, který je spřažen s elektrontransportním řetězcem (čili řadou komplexů, které přenáší elektrony z vyššího energetického stavu na nižší za současného výrazného navýšení redoxního potenciálu) a syntézou ATP z ADP (viz obr. 2).



Obr. 2. Spřažení Krebsova cyklu a syntézy ATP v mitochondriích (podle Alberts et al. 1998). ETC=elektron transportní řetězec.

Kdyby nebylo oxidativní fosforylace, eukaryotické buňky by byly, co se získávání energie týká, odkázány pouze na anaerobní glykolýzu (neboli fermentaci). Ta je však poměrně málo efektivní – z jedné molekuly glukózy při ní buňka získá jen 2 molekuly ATP. Oproti tomu při oxidativní fosforylaci je zisk 38 molekul ATP na jednu molekulu glukózy. Tento rozdíl je způsoben tím, že produkty fermentace obsahují ještě obrovské množství energie. Na základě těchto informací by se mohlo zdát, že buňka vlastně vůbec nepotřebuje využívat mechanismu

anaerobní glykolýzy. Přesto se však tyto metabolické dráhy u eukaryot uchovaly – je tomu tak proto, že anaerobní glykolýza je výrazně rychlejší než aerobní respirace, a je tedy využívána tehdy, když buňka potřebuje generovat velké množství ATP ve velmi krátkém čase (což může být výhodné při krátkodobě zvýšené svalové zátěži – anaerobní glykolýza rychle poskytne dostatečné množství energie například při rychlém útěku či pronásledování kořisti).

Pro lepší pochopení možného významu záměn v genech mtDNA se na elektron transportní řetězec a funkci jednotlivých účastníků se proteinů (z nichž některé jsou kódované v mtDNA) podíváme trochu podrobněji.

Elektron transportní (někdy také *dýchací*) řetězec v mitochondriích je tvořen více než 80 proteiny, přičemž samotného přenosu elektronů se účastní asi 15 z nich (Alberts et al. 1998). Tyto proteiny dohromady vytváří pět velkých enzymových komplexů, které jsou zanořeny a ukotveny ve vnitřní membráně mitochondrií transmembránovými proteiny:

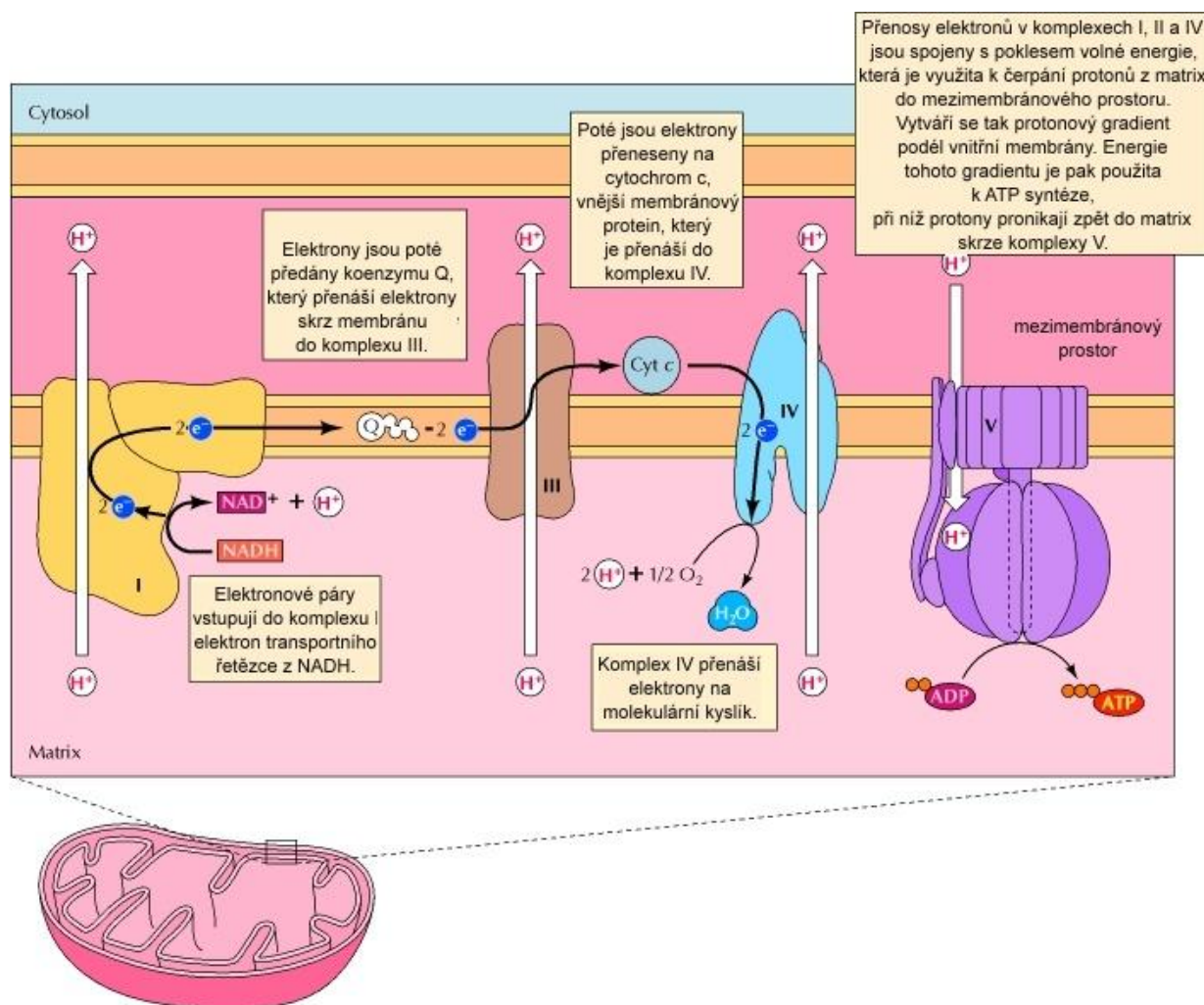
- I. NADH-dehydrogenázový komplex,
- II. sukcinát dehydrogenázový komplex,
- III. komplex cytochromů b-c₁,
- IV. cytochromoxidázový komplex, a
- V. ATP-syntázový komplex.

V membráně je také přítomen koenzym Q (neboli ubichinon). Ubichinon je malá hydrofobní molekula (jako jediný přenašeč elektronů není součástí žádného proteinového komplexu), která se může poměrně volně pohybovat v lipidové dvojvrstvě vnitřní membrány. Reguluje přenos elektronů z komplexu I a II na komplex III.

Samotný elektron transportní řetězec pak vypadá následovně:

Elektrony z NADH vstoupí do elektron transportního řetězce přes komplex I. Zde jsou nejprve přeneseny z NADH na flavin mononukleotid (FMN) a dále přes Fe-S protein na ubichinon. Tímto přenosem se uvolní energie $\Delta G^{\circ'} = -16.6$ kcal/mol. Následně jsou elektrony přeneseny na komplex cytochromů b a c₁ odsud na cytochrom c, čímž se uvolní $\Delta G^{\circ'} = -10.1$ kcal/mol. Cytochrom c, který je vázán ve vnější části vnitřní membrány, pak přenáší elektrony na komplex cytochromoxidáz, kde jsou přeneseny na O₂ ($\Delta G^{\circ'} = -25.8$ kcal/mol). S každým přenosem elektronu a poklesem volné energie dochází zároveň k velkému nárůstu redoxního

potenciálu (v matrix se kumulují H^+), kterého pak využívají ATP-syntázy. Pro lepší představu viz obr. 3.



Obr. 3. Elektron transportní řetězec v mitochondriích. Podle Coopera 2000.

Pro syntézu ATP jsou však důležité pouze komplexy I, III, IV a V – při přenosu elektronu z komplexu II na koenzym Q totiž nedochází k významnému poklesu volné energie, a tím pádem tento proces není sprzęžen se syntézou ATP (Cooper, 2000).

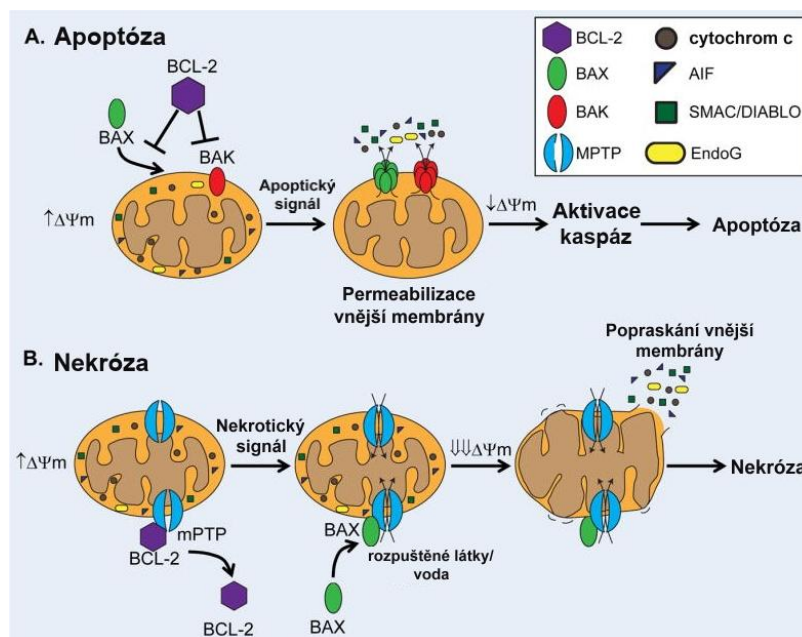
Proteiny, které vytváří jednotlivé komplexy dýchacího řetězce, jsou kódovány jak v jaderné, tak mitochondriální DNA. Většina mitochondriálních proteinů, tedy ~1000-2000, záleží na typu tkáně (Stojanovski et al. 2003), je kódovaná v jádře, a pouze 13 proteinů, které se účastní dýchacího řetězce, je kódováno mtDNA. Má práce se však soustředit na variabilitu na úrovni mtDNA a možný vliv mutací v genech mtDNA na fenotyp jedince a potažmo jeho adaptivní význam. Proto níže uvádím tabulku (tab. 1), která ukazuje, jakou funkci mají v oxidativní fosforylaci jednotlivé proteinové komplexy, a které proteiny kódované mtDNA

jsou v nich přítomny. Lze tak snadno odvodit, která část dýchacího řetězce by mohla být záměnou v daném genu mtDNA ovlivněna.

Tab. 1 - Komplexy figurující v oxidativní fosforylaci.

| komplex | funkce | počet proteinů v komplexu | proteiny kódované mtDNA |
|---|---|--------------------------------------|------------------------------------|
| I - NADH-dehydrogenázový komplex | přenos elektronů z NADH na flavin mononukleotid (FMN) | 44 | ND1, ND2, ND3, ND4, ND4L, ND5, ND6 |
| II - sukcinát dehydrogenázový komplex | přenos elektronů ze sukcinátu na FADH ₂ | 4 | - |
| III - komplex cytochromů b-c ₁ | přenos elektronů z <i>cob</i> na Cyt c | 10 | COB |
| IV - cytochromoxidázový komplex | přenos elektronů na finální akceptor - O ₂ | 19 | COX1, COX2, COX3 |
| V - ATP-syntázový komplex | H ⁺ pumpa; syntéza ATP z ADP | 19 | ATPázové podjednotky 6 a 8 |

Dnes již také víme, že mitochondrie hrají významnou roli také v regulaci apoptotických, zánětlivých a nekrotických procesů v buňce (např. Kubli a Gustaffson, 2012, Harijith et al. 2014; viz obr. 4) a rovněž v buněčné signalizaci, kde se jako signální molekuly uplatňují ROS (podrobně viz např. Tait a Green, 2012). Klíčovým procesem je u všech těchto signalizačních drah permeabilizace mitochondriální membrány, jež má za následek uvolnění signálních molekul – cytochromu c a ROS – z mitochondriální matrix do cytosolu a aktivaci příslušných kaspáz. Zásadní roli v regulaci tohoto procesu pak mají především proteiny z rodiny BCL-2, především anti-apoptický protein BCL-2, který reguluje vazbu pro-apoptických proteinů BAK a BAX na vnější mitochondriální membránu.



Obr. 4. Nekrotické a apoptické procesy na membráně mitochondrií a role proteinů BCL-2, BAK a BAX; MPTP = mitochondriální pór (mitochondrial permeability transition pore), endoG = endonukleáza G, AIF = faktor indukující apoptózu (apoptose inducing factor). Podle Kubli a Gustafsson (2012).

3.3.6. Původ mitochondrií

Myšlenkou, že mitochondrie by mohly být bakteriálního původu, se zabývali vědci již krátce poté, co byly tyto buněčné organely objeveny. Více či méně přímé úvahy na toto téma najdeme v mnoha vědeckých pracích – např. Altmana 1890, Mereschovskyho 1910, Portiera 1918 a dalších. Navzdory tomu se však vědci otázce původu mitochondrií začali více věnovat až poté, co bylo zjištěno, že mitochondrie mají vlastní DNA (Nass a Nass 1963) – tento objev jen podpořil dřívější úvahy o tom, že mitochondrie byly původně zcela autonomní bakterie. V roce 1970 Lynn Margulis vydala knihu *Origin of Eukaryotic Cells* (Margulis, 1970), která znovu rozpoutala diskusi; během následujících let bylo navrženo mnoho modelů vysvětlujících endosymbiotický původ mitochondrií (pěkný přehled těchto modelů zpracovali Martin et al. 2001).

Zásadní pro objasnění této otázky byl však především rozvoj sekvenačních a klonovacích technik na přelomu sedmdesátých a osmdesátých let, díky němuž mohly být kompletně osekvenovány mitochondriální genomy člověka (Anderson et al. 1981), myši (Bibb et al. 1981) a skotu (Anderson et al. 1982), čímž se debata posunula do úplně jiné roviny – na úroveň příbuznosti DNA a genů.

Zmíněné práce zároveň stanovily základní strukturu savčí mtDNA – 13 protein kódujících genů, přibližně 20 tRNA genů a dva geny kódující ribozomální podjednotky (rRNA). Pozdější práce na nesavcích modelech ukázaly, že mtDNA je, co se týče délky i genového složení, velice variabilní (mtDNA *Plasmodium falciparum* má například jen asi 6 kbp, oproti tomu mitogenom velké většiny rostlin přesahuje 200 kbp), nicméně má některé společné jmenovatele, a jedním z nich jsou právě výše zmíněné rRNA. Na ně se zaměřila pozornost při zkoumání fylogenetického původu mitochondrií. Malá a velká ribozomální podjednotka si totiž – navzdory velké variabilitě – napříč eukaryoty zachovaly určitý stupeň podobnosti v sekvenci a sekundární struktuře s bakteriemi. Proto byly na základě sekvencí těchto genů sestaveny první fylogenetické stromy (Gray et al. 1984, Cedergren et al. 1988), které ukazovaly na to, že prvotní mitochondrie pochází z domény *Bacteria*.

Významným krokem ve zkoumání původu mitochondrií byly také analýzy rostlinných mitochondriálních rRNA, které jsou zvláště konzervativní a velmi podobně bakteriím (Schnare a Gray 1982). Právě analýzy rostlinných rRNA a jejich srovnání s rRNA bakterií vedly nakonec k závěru, že mitochondrie jsou původem z třídy alfa-proteobakterií (Yang et al. 1985). Další analýzy tuto myšlenku jen potvrdily a označily za nejpravděpodobnějšího předka mitochondrií parazitické bakterie z řádu *Rickettsia* (Andersson et al. 1998).

Dnes již jsme si tedy poměrně jistí tím, že mitochondrie byly původně parazitické bakterie a že vznikly v historii pouze jednou (Gray et al. 1999). Otázka, jakým způsobem probíhalo oddělení mitochondrií od rickettsií, však zůstává nezodpovězena; Gray (2014) navrhuje pre-endosymbiotický model, kdy nejprve v buňce vznikla endogenní organela (premitochondrie) a teprve později došlo ke spojení této organely s α -proteobakterií.

3.4. Modelový organismus – norník rudý (*Clethrionomys glareolus*)

Norník rudý (*Clethrionomys glareolus*, někdy též *Myodes glareolus*), viz obr. 5, je malý hlodavec (průměrně asi 25 g) z čeledi hrabošovitých (*Muridae*). Spadá do jedné podčeledi (*Arvicolinae*) s hraboši (v české literatuře můžeme dokonce najít odkaz na tuto podobnost – dříve býval označován jako hraboš lesní či hraboš rudý). Příbuznost hraboše a norníka je evidentní na první pohled – oba totiž mají, na rozdíl od myši, mohutnější tělesnou konstituci, menší uši a kratší ocas. Norník rudý se od ostatních zástupců hrabošovitých odlišuje především výrazným rudým zbarvením srsti, které je patrné hlavně na hřbetě; srst na břiše a nohách

je bílá. Ve srovnání s ostatními našimi hraboši má také delší ocas, který dosahuje asi poloviny délky těla norníka.

Norník je ve fylogeografii využíván coby zástupce temperátních druhů (čili druhů mírného pásma) v Evropě. Obývá především listnaté a smíšené lesy, ale ve Skandinávii se vyskytuje též



v lesech boreálních. Setkat se s ním můžeme také v polních remízcích, v křovinách poblíž luk, ale též v parcích. Ve Velké Británii často obývá živé ploty podél cest, tzv. *hedges* nebo *hedgerows*, často pěstované na valech zeminy označovaných jako *hedge banks* – odtud anglický název norníka *bank vole*. Živí se hlavně semeny, bobulemi, kořínky, mladými větvičkami rostlin, houbami a ořechy, ale také drobným hmyzem.

Areál rozšíření norníka sahá od oblasti Středozemního moře až za polární kruh, vyskytuje se od nejzápadnějších oblastí Evropy až k Altaji na Sibiři (viz obr. 6, oblast výskytu je zabarvena oranžově). Právě výskyt po takřka celé Evropě a obývání klimaticky odlišných stanovišť dělají z norníka skvělý modelový organismus, který je v poslední době využíván především ve fylogeografických studiích zabývajících se postglaciální kolonizací Evropy.

3.4.1. Fylogeografie norníka

Většina prací, které byly dosud zveřejněny o fylogenezi a fylogeografii norníka, využívala jako marker cytochrom b (*cob*), velmi populární marker při studiu vnitrodruhové i mezidruhové variability nejen savců (Johns a Avise, 1998). Na základě těchto dat pak bylo v Evropě popsáno osm základních mtDNA linií. Pomineme-li fosilní nálezy, jsou právě analýzy geografického rozšíření mtDNA linií a jejich vzájemné příbuznosti hlavním nástrojem při hledání odpovědi na otázku, jak probíhala kolonizace Evropy temperátními druhy v průběhu změny klimatu, která následovala po poslední době ledové (jejíž vrchol se označuje *Last Glacial Maximum*, LGM).

3.4.1.1. Mitochondriální linie norníka v Evropě

Na základě dat z *cob* bylo v Evropě popsáno osm, respektive devět (počítáme-li linii uralskou, jejíž haplotypy odpovídají norníku tajgovému, *Clethrionomys rutilus*; viz Tegelström 1987) mitochondriálních linií norníků, viz obr. 6. Při jejich popisu budu vycházet z terminologie poprvé použité v člancích Deffontaine et al. (2005) a Kotlíka et al. (2006).

Západní linie má velmi široký areál rozšíření – od severní části Balkánského poloostrova přes střední Evropu až na úplný západ Evropy (včetně Velké Británie).

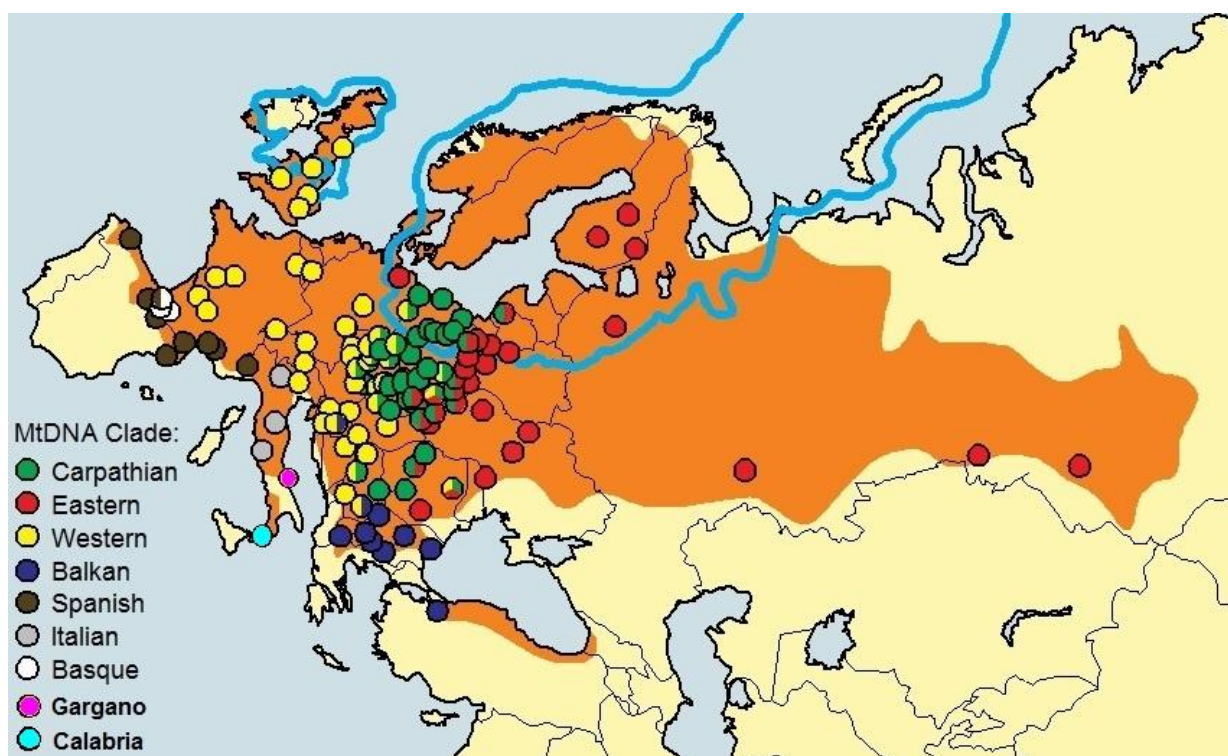
Východní linie se pak vyskytuje od východní Evropy (Rumunsko, Ukrajina) směrem na sever (Finsko) a na východ až na Sibiř, ale setkáme se s ní i v severnější položených zemích ve střední Evropě (severní část Německa, Dánsko).

Karpatskou linii, popsanou Kotlíkem et al. (2006), nejčastěji nacházíme ve střední Evropě v oblasti karpatského oblouku a odtud směrem na sever až k Baltskému moři (Polsko).

Balkánskou linii pak najdeme na Balkánském poloostrově a v severním Turecku na pobřeží Černého moře.

Na jihu Francie a severu Španělska v oblasti Pyrenejí se vyskytují především jedinci původem ze španělské linie, nicméně v Baskicku, malé oblasti na pomezí Francie a Španělska, která se od těchto zemí značně odlišuje svou kulturou a jazykem, byla nedávno popsána ještě jedna linie – baskická (Deffontaine et al. 2009). Tato linie se při fylogenetických analýzách *cob* jeví jako bazální ke všem ostatním liniím. Překvapivý výsledek přinesla práce Collangelo et al. (2012), v níž při tvorbě fylogenetických stromů jeden ze vzorků z Itálie klastroval do stejného kládu s baskickými. Jako jedno z možných vysvětlení se nabízí, že baskická linie je ve skutečnosti soubor jaderných pseudogenů a zmíněný vzorek z Itálie taktéž. Mezi baskickým regionem a Itálií je poměrně velké množství geografických překážek a výskyt jedince s mtDNA baskické linie v Itálii je tak velice nepravděpodobný.

Na jihu Evropy, na Apeninském poloostrově, pak nacházíme dvě mitochondriální linie, italskou a v oblasti Kalábrie na nejjižnější výběžku Apeninského poloostrova kalabrijskou, která patří mezi evolučně nejstarší z dosud popsaných linií (Colangelo et al. 2012). Zdá se rovněž, že existuje ještě další velmi odlišná italská linie v oblasti Gargano (Colangelo et al. 2012). Pro lepší přehled viz obr. 6 (upraveno podle Wójcika et al. 2010).

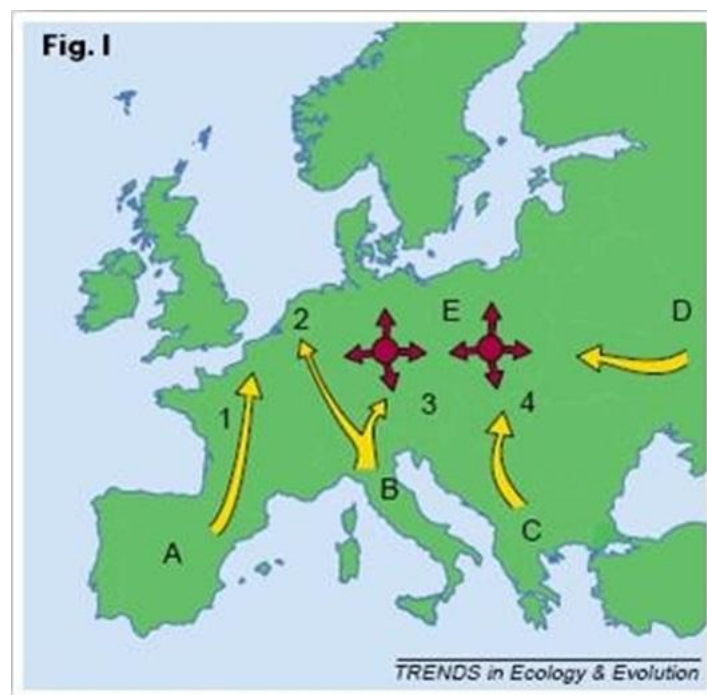


Obr. 6 Rozšíření norníka rudého a výskyt jednotlivých linií mtDNA (upraveno podle Wójcika et al. 2010).

3.4.1.2. Norník jako klíčový model postglaciální kolonizace Evropy temperátními druhy

Klima v Evropě se během posledních 2 milionů let mnohokrát dramaticky změnilo – vystřídalo se několik dob ledových s dobami meziledovými. Tyto změny klimatu byly spojeny se změnou plochy, kterou v Evropě pokrýval pevninský ledovec. Právě rozšíření ledovce během pozdního pleistocénu (s maximem cca 23 000 – 18 000 AC) mělo zásadní vliv na rozšíření temperátních druhů. Klima ve střední Evropě bylo v období LGM suché a studené a podobalo se spíše klimatu, se kterým se setkáváme v tundrách. Proto se druhy mírného pásma stáhly do jižnějších částí Evropy, tzv. refugií, což byla místa s méně extrémními klimatickými podmínkami.

Alespoň donedávna se předpokládalo, že refugia temperátních druhů se v Evropě nacházela výhradně na jihu na třech středomořských poloostrovech – Balkánském, Apeninském a Pyrenejském (Taberlet et al. 1998) - a odsud se po ústupu ledovce tyto druhy znovu rozšířily po Evropě (detailní shrnutí např. Hewitt 2004). S postupem času se však objevily jak nové fosilie, tak názory, které tuto myšlenku rozporují (Bilton et al. 1998, Stewart a Lister, 2001; Deffontaine et al. 2005, Kotlík et al. 2006, Sommer a Nadachowski 2006, Wójcik et al. 2010) a navrhuji ještě jiné geografické umístění refugií (viz obr. 7).



Obr. 7. Refugia temperátních druhů během LGM a předpokládaný směr rekolonizace Evropy po ústupu zalednění (Steward a Lister, 2001). Jižní refugia jsou zaznačena žlutě, severní refugia (jejichž objev znamenal zásadní změnu pro studium postglaciální kolonizace Evropy) červeně.

Norník rudý je jedním z druhů, u nichž se začala uvažovat hypotéza severních refugií², a na základě molekulárních dat (Bilton et al. 1998, Deffontaine et al. 2005) a fosilních nálezů (např. Horáček 2000) pak jako možná oblast, v níž tento druh přežíval, byly navrženy Karpaty. Myšlenka, že jedno z možných severních refugií norníka se nacházelo v Karpatech, byla podpořena Kotlíkem et al. (2006), kteří v této oblasti objevili novou mitochondriální linii, zcela odlišnou od ostatních linií dosud popsanych v Evropě, která navrženému scénáři odpovídala i svým evolučním stářím. Tuto hypotézu dále zkoumali Searle et al. (2009) a svou pozornost zaměřili na Británii, která je z hlediska studia postglaciální kolonizace skvělým modelovým územím.

Británie byla totiž v době, kdy pevninský ledovec v Evropě začal ustupovat (tedy cca 19 000 AC), ještě spojena se zbytkem Evropy pevninským mostem. Po tomto mostu pak byla postupně kolonizována – v případě norníka ve dvou vlnách. První norníci do Británie pravděpodobně dorazili brzy poté, co se po ústupu pevninského ledovce zmírnilo klima. O asi 2 000 let později, po přechodném návratu chladného klimatu (období označované jako Mladší dryas), pak následovala druhá vlna kolonizace (Searle et al. 2009). S postupujícím globálním oteplováním

² Termín „severní refugia“, který budu nadále používat, pochází z článku Stewarta a Listera (2001) a vztahuje se k refugiím položeným severněji v Evropě, než ve Středomoří.

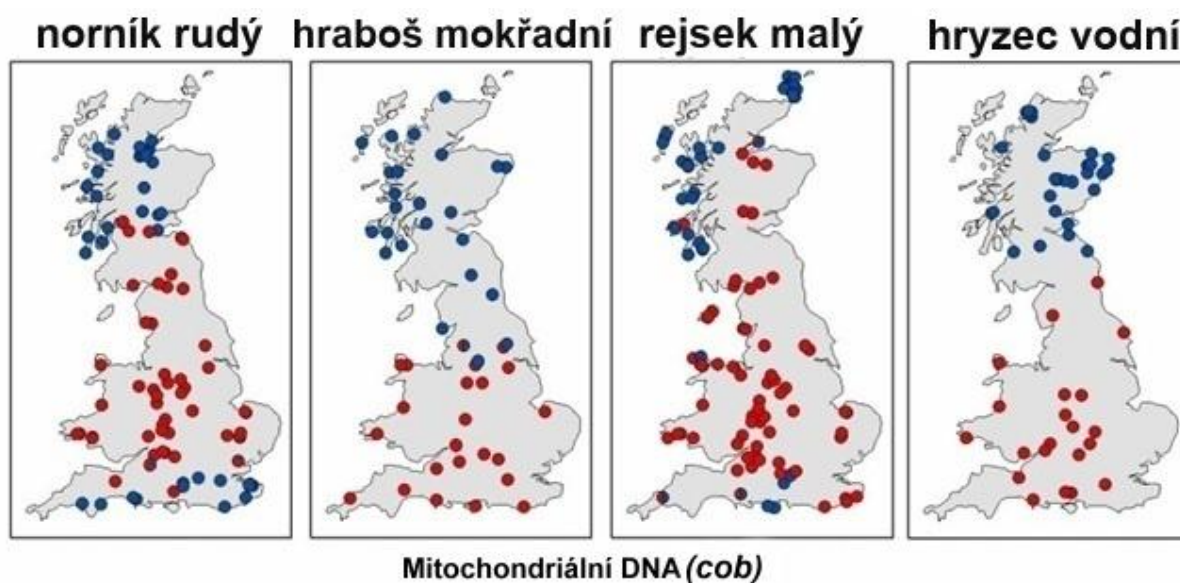
se zvyšovala i hladina moře a cca 8000 AC byla oblast, která původně spojovala Evropu a Británii (tzv. Doggerland), zatopena a Británie oddělena vodami dnešního Severního moře. Změna klimatu tak měla za následek vznik ostrovní populace, v níž se vyskytují dvě mtDNA linie, které během posledních 10 000 let nebyly ovlivněny genovým tokem z populací na kontinentu.

3.4.1.3. Linie mtDNA norníků v Británii

Co je však na těchto dvou mtDNA liniích nejzajímavější, je jejich geografická distribuce. V Británii totiž u norníka, stejně jako u některých dalších druhů hlodavců a hmyzožravců, můžeme pozorovat tzv. Keltský lem (*Celtic fringe*). Tento výraz byl původně použit k popisu současné geografické distribuce lidských populací v Británii. „Keltský lem“ je zde tvořen populacemi, které jsou potomky prvních kolonistů, kteří přišli do Británie po LGM, tedy Keltů³. Tyto „keltské“ populace se dnes nacházejí na severním a západním okraji Británie (Wales, Skotsko, Cornwall) a v severní Francii (oblast Bretaně) a od ostatních populací v Británii se odlišují jak z hlediska genetického, tak z hlediska lingvistického. Dnešní geografická distribuce těchto populací v Británii do jisté míry odráží, jakým způsobem byla Británie kolonizována po LGM, tedy ve dvou vlnách - jako první přišly do Británie keltské kmeny, po nich anglosaské.

Podobný Keltský lem, kdy populace při severním, západním a/nebo jižním okraji ostrova mají specifické genetické charakteristiky, můžeme pozorovat u několika druhů drobných savců v Británii. Searle et al. (2009) zaznamenali podobný distribuční vzor mtDNA (*cob*) u 3 dalších druhů (viz obr. 8) – kromě norníka rudého také u hraboše mokřadního (*Microtus agrestis*), rejška malého (*Sorex minutus*) a hryzce vodního (*Arvicola terrestris*; data z práce Pierniey et al. 2005).

³ Zde se jedná o jiné keltské kmeny než ty, které obývaly střední Evropu; více viz Searle et al. 2009.



Obr. 8. Tzv. Keltský lem u různých druhů malých savců v Británii (podle Searle et al. 2009). U všech je vidět jistý severojižní trend, kdy je linie prvních kolonistů vytlačována linií, jež přišla do Británie jako druhá.

Zajímavé jsou na tomto geografickém rozložení populací dvě věci. Zaprvé, mezi jednotlivými populacemi mtDNA jsou velice ostré hranice rozšíření; ve vzorkovaných lokalitách se většinou vyskytovali vždy pouze jedinci z jedné mtDNA linie. Tuto distribuci nelze jednoduše vysvětlit žádnými fyzicko-geografickými bariérami, které by se vyskytovaly v oblastech na rozhraní těchto populací. Zadruhé, u všech zkoumaných druhů se objevuje severojižní trend, kdy v severněji položených lokalitách (Skotsko) nacházíme téměř výhradně jedince nesoucí mtDNA prvních kolonistů.

Na severu Británie je výrazně drsnější a studenější klima než ve středních a jižních částech Británie. Nabízí se otázka, zda geografické rozšíření těchto dvou mtDNA v Británii není důsledkem selekční výhody jedné linie nad druhou. Případ selekční výhody ve vyšších nadmořských výškách spojené s určitým haplotypem mtDNA byl popsán např. v předhůří Alp u bělozubky tmavé (*Crocidura russula*; Fontanillas et al. 2005), hmyzožravce z čeledi rejskovitých.

První populace norníků, která kolonizovala Británii, přišla pravděpodobně ze severněji položených refugií než druzí kolonisté (Searle et al. 2009). Vztahy britských linií mtDNA norníka s liniemi známými z kontinentu však dosud podrobně studovány nebyly. Je možné předpokládat, že klimatické podmínky v severních refugiích byly drsnější než v refugiích jižních. Je tedy možné, že norníci ze severních refugií byli lépe adaptováni na nepříznivé podmínky, a že tím pádem měli výhodu v počáteční fázi postglaciální kolonizace Evropy

v době, kdy klima sice bylo mírnější než v LGM, ale pro temperátní druhy stále poměrně extrémní. Není vyloučeno, že důvodem, proč dnes obývají spíše hornaté a chladnější oblasti Británie, je právě adaptace na tyto horší podmínky. Fakt, že druzí kolonisté (nebo alespoň jejich mtDNA) na velkém území Anglie nahradily ty první, zase naznačuje možnost, že později příchozí populace mohla mít v době svého příchodu nějakou selekční výhodu oproti té první. Kandidátních genů, které by mohly hrát roli v adaptaci organismu na nepříznivé klimatické podmínky nebo v soutěži mezi zástupci různých populací, je nespočet. Jako jedna z možností se nabízí například hemoglobin (Kotlík et al. 2014) – přenašeč kyslíku, který je potřebný coby finální akceptor elektronu v dýchacím řetězci. U hemoglobinu norníka byla totiž popsána velmi podobná geografická distribuce v Británii jako u mtDNA (Hall 1979; Kotlík et al. 2014).

Z potenciálních cílů adaptivní evoluce jsem se ve své studii zaměřila na geny mtDNA kódující podjednotky proteinových komplexů dýchacího řetězce (význam těchto mitochondriálních proteinů podrobněji zmiňuji v kapitole o mitochondriích).

Tento výběr není náhodný. Jednak byly už u mnoha druhů popsány případy adaptivní evoluce mtDNA (Nachman et al. 1996, McClean et al. 2004, Foote et al. 2011) a jak jsem zmínila výše, argumenty pro možný výskyt adaptivní evoluce mtDNA existují také u norníka. Především byl ale Keltský lem u norníků v Británii popsán právě na základě mtDNA a není zřejmé, kolik jaderných genů vykazuje podobný vzor. Je proto logickým prvním krokem hledat možnou příčinu této distribuce (funkční rozdíly) právě v mtDNA. Rovněž dosud chybělo ověření existence a spolehlivá rekonstrukce evolučních vztahů linií mtDNA v Evropě sekvencemi kompletních mitogenomů a přiřazení dvou linií norníka, vyskytujících se v Británii, k těmto kontinentálním liniím.

4. Materiál a metody

4.1. Výběr jedinců ke studii

Mým cílem bylo získat sekvence kompletní mtDNA minimálně dvou jedinců od každé linie. To však nebylo vždy úplně možné, protože pro linii španělskou a garganskou nebyl k dispozici dostatečný počet vzorků v požadované kvalitě. Jako outgroup, pomocí něhož jsem pak kořenila fylogenetický strom, jsem zvolila jiný druh rodu *Clethrionomys* – norníka tajgového (*C. rutilus*), jehož mtDNA byla v mých vzorcích norníka rudého přítomná v podobě introgrese, tj. uralské linie.

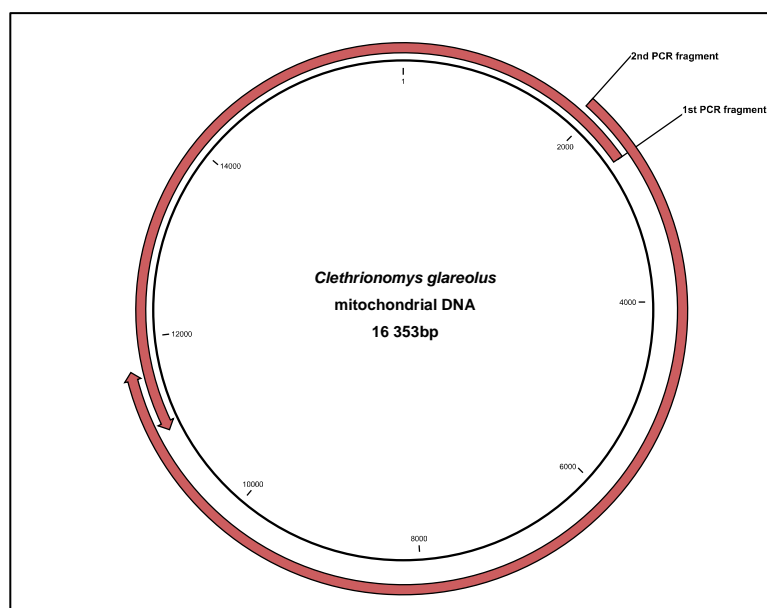
4.2. Izolace DNA

Celkovou DNA z 58 norníků jsem izolovala z tkání, uchovávaných v etanolu. Výchozím materiálem byla nejčastěji tkáň ze sleziny či jater, která jsou pro velký výtěžek mtDNA optimální, v některých případech ale musely stačit i prsty nebo kousky ucha. K izolaci jsem použila DNeasy Blood & Tissue Isolation Kit od firmy Qiagen (Valencia, CA, USA), přičemž jsem postupovala podle protokolu doporučeného výrobcem.

4.3. Amplifikace mtDNA – long-range PCR

MtDNA jsem získala z celkové DNA pomocí PCR amplifikace dlouhých úseků (dále jen long-range PCR).

Při PCR kratších úseků na mtDNA totiž není ničím výjimečným současná amplifikace nukleárních mitochondriálních pseudogenů (tzv. *numts*, viz výše). Současná amplifikace úseků mtDNA a *numts* může způsobit značné problémy, a to jak při sekvenování (dvojité píky nebo až celková nečitelnost chromatogramu), tak při fylogenetické analýze (*numts* se totiž, díky nižší mutační rychlosti v jaderném genomu, mohou při analýzách jevit jako „ancestrální“ sekvence, a mohou vést ke vzniku velmi matoucích fylogenetických stromů). Abych minimalizovala riziko amplifikace *numts*, využila jsem kruhového tvaru molekuly mtDNA a použila metodu long-range PCR k amplifikaci mtDNA pomocí dvou překrývajících se dlouhých fragmentů (schematicky zobrazeno na obr. 9).



Obr. 9. Schéma amplifikace mtDNA pomocí dvou překrývajících se PCR fragmentů.

4.3.1. Primery

Primery pro long-range PCR byly navrženy do konzervativních oblastí v mtDNA (*rrnL*, *nd4*, *trnL1*, *trnS1*), které byly vybrány na základě alignmentu známých sekvencí mtDNA hlodavců - myši středoevropské *Mus musculus musculus* (NC_010339), hraboše tchajwanského *Microtus kikuchii* (NC_003041), hraboše východoevropského *Microtus levis* (NC_008064), myši *Mus terricolor* (NC_010650), myšice Chevriery *Apodemus chevrieri* (HQ896683), myšice temnopásé *Apodemus agrarius* (JN629047), krysy obecné *Rattus rattus* (NC_012374), norníka sečuánského *Eothenomys chinensis* (NC_013571), křečika čínského *Cricetulus griseus* (NC_007936), křečka zlatého *Mesocricetus auratus* (NC_013276) a korejského norníka *Clethrionomys regulus* (NC_016427).

Nejprve byly použity dříve publikované 2 páry primerů, jimiž byla původně získána sekvence korejského norníka *C. regulus* (Kim a Park 2012). Protože však tyto primery nedávaly u norníka rudého dobré výsledky, bylo pro amplifikaci navrženo dalších 18 primerů. Celkem bylo vyzkoušeno 8 párů amplifikačních primerů pro první fragment a 10 pro druhý fragment.

Pro amplifikaci prvního fragmentu byly nakonec vybrány primery 11037F21 (forward) a 2477R20 (reverse), druhý fragment byl získán pomocí primerů 1900F21 (forward) a 11658R21 (reverse).

Sekvence všech použitých amplifikačních primerů uvádím v tabulce P2 v přílohách.

4.3.2. Použité reakční směsi a polymerázy

Na long-range PCR jsem vyzkoušela dvě různé polymerázové směsi – LA DNA Polymerase (Top-Bio, Praha 4, ČR) a LongAmp Taq DNA Polymerase (New England Biolabs, Ipswich, MA, USA).

První fragment pro prvního jedince (1634) jsem získala amplifikací LA DNA polymerázou. Na druhý fragment jsem použila jinou polymerázu - LongAmp Taq DNA Polymerase. Vzhledem k výborným a vyrovnaným výsledkům, kterých jsem předtím s LA DNA polymerázou nedosahovala, jsem se rozhodla ji použít při amplifikaci obou fragmentů mtDNA u všech dalších vzorků. Výhodou LongAmp Taq DNA Polymerase je, že syntetizuje již při 65°C – nižší teplota extenze má za následek pomalejší tepelnou degradaci polymerázy, což lze u dlouhých úseků poznat především na vyšším výsledném množství PCR produktu.

4.3.2.1. LA DNA Polymerase (Top-Bio)

PCR probíhala v 25 µl reakční směsi. Výsledné koncentrace složek směsi byly tyto: 1× LA PCR reakční pufr bez MgCl₂, 2.5 mM MgCl₂, 0.5 mM dNTPs, 0,5 µM (forward i reverse) primerů, 1.25 u LA DNA polymerázy, ~10 ng/µl DNA.

4.3.2.2. LongAmp Taq DNA Polymerase (New England Biolabs)

PCR jsem prováděla v 25 µl reakční směsi, která odpovídala protokolu dodanému výrobcem. Výsledná koncentrace složek PCR směsi byla následující: LongAmp Taq Reaction Buffer (1×), 300 µM dNTPs, 0.4 µM (forward i reverse) primerů, 2.5 u LongAmp Taq DNA polymerázy a ~10 ng/µl DNA v 25 µl PCR směsi.

4.3.3. PCR programy

Veškeré PCR probíhaly na přístrojích C1000TM Thermal Cycler a DNA Engine® Peltier Thermal Cycler od firmy Bio-Rad Laboratories.

4.3.3.1. LA DNA Polymerase

Při použití LA DNA polymerázy jsem používala 2 různé programy:

1. fragment (~7800 bp)

Počáteční denaturace – 94°C/ 2 min, denaturační krok – 94°C/ 30 s, annealing – 54°C/ 40 s, extenze – 72°C/ 8 min, cyklus 34× opakován, finální extenze 72°C/ 10 min.

2. fragment (~9500 bp)

Počáteční denaturace – 94°C/ 2 min, denaturační krok – 94°C/ 30 s, annealing – 54°C/ 40 s, extenze – 72°C/ 10 min, cyklus 34× opakován, finální extenze 72°C/ 10 min.

4.3.3.2. LongAmp Taq DNA Polymerase

Pro amplifikaci obou fragmentů byl použit stejný program:

Počáteční denaturace - 94°C/ 30 s, denaturační krok - 94°C/ 20 s, annealing – 54°C/ 30 s, extenze – 65°C/ 9 min, cyklus 30× opakován, finální extenze – 65°C/ 10 min.

4.4. Gelová elektroforéza

Kontrola čistoty produktů a výtěžnosti PCR byla provedena pomocí elektroforetické verifikace.

Agarózový gel byl připraven rozmícháním agarózy (SERVA, kvalita *research grade*) v 0,5% TBE pufru a krátkým povařením v mikrovlnné troubě. Výsledná koncentrace agarózy byla 0,8 %. Gel byl obarven GoldViewTM (Viswagen Biotech).

Gelová elektroforéza probíhala v TBE pufru při pokojové teplotě na přístroji Labnet Power Station 300. Prvních 20 minut probíhalo při napětí 80 V, dalších 40 minut potom při 98 V. Se vzorky byl na gel nanesen také velikostní marker (GeneRuler 1kb Ladder, Thermo Scientific).

Výsledky gelové elektroforézy byly následně zkontrolovány a nafoceny pomocí transluminátoru MultiImage™ Light Cabinet (Alpha Innotech Corporation).

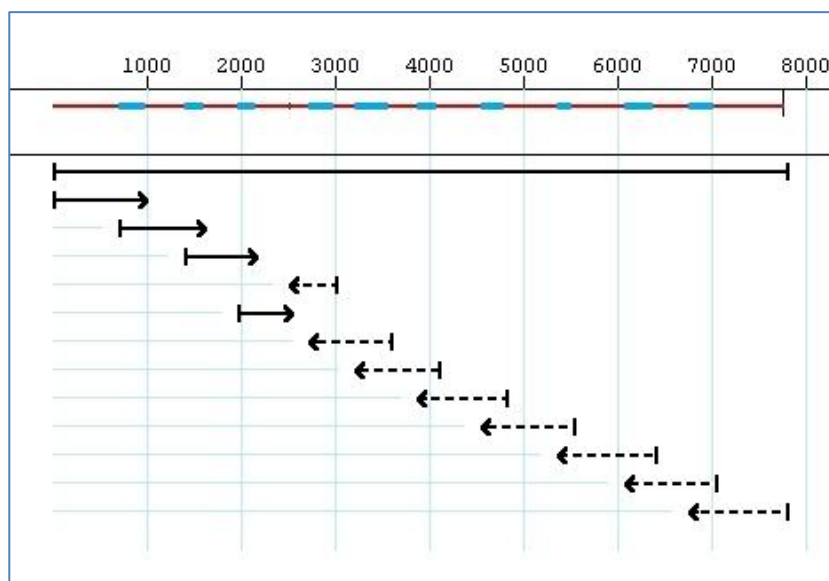
4.5. Purifikace PCR produktů

Získané amplikony byly purifikovány pomocí kitu (QIAquick PCR Purification Kit, Qiagen). Můj postup se od standardního protokolu lišil pouze v posledním kroku, kdy jsem eluci produktu místo běžných 1×50 µl elučního pufru EB prováděla ve dvou krocích, tedy 2×25 µl elučního pufru EB (pro vyšší výtěžek DNA při tomto postupu).

Všechny purifikace byly prováděny na centrifuze Mikro 200 firmy Hettich Zentrifugen.

4.6. Primer walking – první referenční sekvence

Pro získání referenční sekvence⁴ normíka rudého jsem zvolila sekvenační strategii procházení primerem (primer walking). Schéma pro primer walking u prvního fragmentu viz obr. 10.



Obr. 10. Schéma primer walking sekvenace prvního PCR fragmentu (~7800 bp). Forwardní primery jsou vyznačeny plnou čarou, reverzní přerušovanou čarou. Sekvence primerů viz tabulka P2 v přílohách.

Celkem jsem pro primer walking použila 24 nově navržených sekvenačních primerů, využila jsem také již navržených amplifikačních primerů (viz seznam primerů v tabulce P2 v přílohách). První fragment byl sestaven z 12 navzájem se překrývajících sekvencí, druhý fragment potom z 15 takových sekvencí.

⁴ referenční sekvence pro sestavení mitogenomů z dat RNA-Seq

Získané sekvence jsem seřadila pomocí programu CodonCode Aligner (verze 5.0.1.6), celý alignment jsem ručně prošla, zkontrolovala a opravila, a pro každý fragment exportovala konsenzuální sekvenci. Tyto exportované konsenzuální sekvence pro oba fragmenty jsem pak ručně spojila a srovnala s referenční sekvencí korejského norníka *C. regulus* v programu BioEdit 7.0.9.0. Výsledná sekvence pak byla anotována na základě anotace genomu myši (NC_010339) a *C. regulus* (NC_016427). Anotace byly ověřeny a některé tRNA geny nově anotovány pomocí programu MITOS (Bernt et al. 2013). Anotovaná první sekvence byla vložena do databáze GenBank (KF918859) a publikována v časopise *Mitochondrial DNA* (Bendová et al., 2014; kompletní článek viz příloha P5).

4.7. Sekvenování metodou Sangerova

První sekvence získaná metodou primer walking byla použita pro návrh primerů pro Sangerovo sekvenování. Celkem bylo při dalším sekvenování použito 40 primerů, přičemž jsem využila 18 primerů použitých dříve na primer walking a také 4 primery, původně navržené pro long-range PCR. Pro potřeby další sekvenace jsem tedy navrhla 18 primerů.

Při návrhu nových sekvenačních primerů jsem se snažila, aby primery byly použitelné pokud možno pro všechny linie. Zcela cíleně jsem zavrhl návrh degenerovaných primerů (z důvodu zachování dostatečné specifity primerů). Při návrhu jsem využívala jednak získané referenční sekvence, která odpovídá karpatské linii mtDNA, a také sekvencí z RNA-Seq dalších jedinců z Británie (ze západní a karpatské linie; viz dále), případně dostupných sekvencí *cob* norníků z databáze GenBank. Prvotní sada primerů (obsahující 28 primerů) výborně fungovala na západní a karpatskou linii, ale pro sekvenaci dalších linií jsem za účelem dosažení co nejlepších výsledků dodatečně navrhla dalších 10 primerů (tj. některé primery byly při sekvenování dalších linií nahrazeny jinými, které dávaly lepší výsledky).

Při sekvenaci uralské linie (mtDNA *C. rutilus*) jsem opět musela navrhnout nové primery (celkem 8), neboť rozdíly oproti sekvenci *C. glareolus* byly takové, že mnoho původních primerů bylo nepoužitelných z důvodu nedostatečné komplementarity.

Většina primerů fungovala velice dobře a poskytla kvalitní sekvence hned při prvním sekvenování. Některé sekvence musely být opraveny; pokud nebylo dosaženo dobrého výsledku ani po druhé opravě a dané místo nebylo pokryto alespoň jednou spolehlivou sekvencí, přistoupila jsem k návrhu nového primeru pokrývajícího chybějící oblast.

Veškeré sekvenační práce byly realizovány firmou Macrogen Inc. (Soul, Jižní Korea), přičemž menší objednávky byly zpracovávány pobočkou v Nizozemsku (Amsterdam) a větší potom pobočkou v Jižní Koreji. Získané sekvence byly vloženy do databáze GenBank (KM892809 – KM892841) při příležitosti publikace výsledků práce formou článku Filipi et al. (2015) v časopise *Mol. Phylogenet. Evol.* (viz příloha P6).

4.8.RNA-Seq norníků z Velké Británie

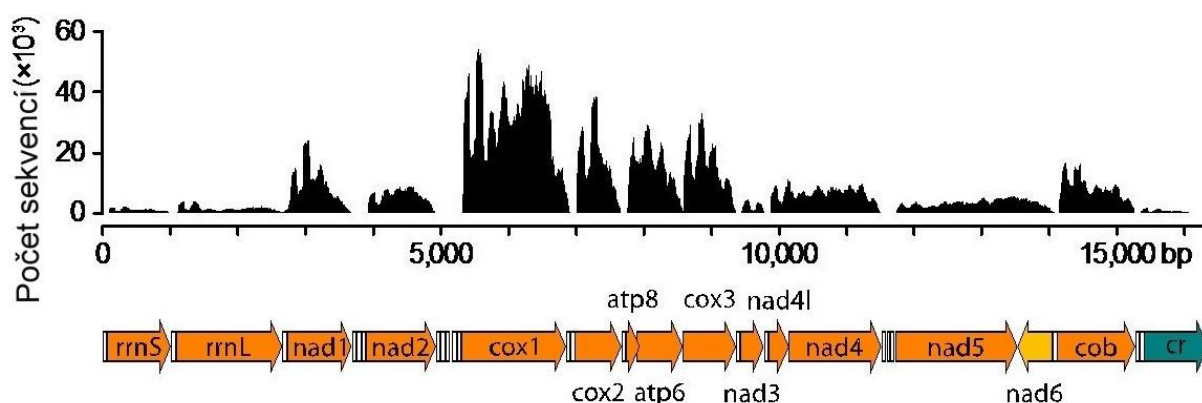
U 3 vzorků z Velké Británie jsem pro získání kompletních sekvencí mitogenomů využila data z NGS sekvenování RNA (RNA-Seq) jedinců z Velké Británie, která jsem měla k dispozici z jiného projektu, který běží v naší laboratoři (viz např. Marková et al. 2014; Kotlík et al. 2014).

RNA-Seq je metoda, která pracuje s RNA jako výchozím materiálem, a je v podstatě spojením reverzní transkripce s následným sekvenováním získané cDNA. Podrobněji pak celý proces sekvenování sestává z přípravy RNA (kdy dochází k obohacení směsi o požadovaný typ RNA; v případě mých vzorků polyA, tedy mRNA), přípravy cDNA knihovny (což zahrnuje fragmentaci a selekci fragmentů RNA, které mají vhodnou délku k sekvenování – zde zvoleno 250–400 bp –, reverzní transkripci do cDNA, připojení adaptérů na konce cDNA, amplifikaci správně vytvořených fragmentů cDNA pomocí PCR a kvantifikaci a kontrolu kvality) a sekvenování (v tomto případě pomocí tzv. modulu pair-end 2×100 bp, tedy 100 bp z každého konce každého fragmentu).

Vzorky byly osekvenovány na přístroji Illumina HiSeq 2000, který pracuje na principu SBS (*Sequencing By Synthesis*). Tento sekvenační systém detekuje připojení fluorescenčně značeného reverzibilního terminátoru, což zajišťuje detekci připojení pouze jedné báze v každém cyklu. Zásadní výhoda spočívá v přesné detekci sekvence i v případě opakujících se bází (což je u některých jiných NGS metod problematické).

Z obrovského množství dat (ca. 50 milionů sekvencí – *reads* – pro každého jedince), která byla díky RNA-Seq získána, jsem použila jen sekvence, které měly sekvenční shodu s mojí referenční sekvencí mitogenomu norníka rudého (KF918859, Bendová et al. 2014). Alignment (*mapping*) byl vytvořen v programu CLC Genomics Workbench 6.0.1. Vybrané sekvence se musely shodovat s referenční sekvencí minimálně v 90 % své délky a musely mít minimálně 95% sekvenční podobnost s referenční sekvencí.

Tento postup využívá toho, že všechny mitochondriální mRNA, rRNA i tRNA jsou získávány sestřihem z několika málo (u člověka dvou) dlouhých polycistronních prekurzorových transkriptů, které pokrývají téměř celou molekulu mtDNA (Temperley et al. 2010). Sekvence získané z RNA-Seq však přesto bohužel neposkytly dostatečně rovnoměrné pokrytí pro získání zcela kompletní sekvenční mtDNA (viz obr. 11).



Obr. 11. Pokrytí jednotlivých oblastí mitogenomu metodou RNA-Seq. Na ose x je pozice na mitogenomu, osa y pak ukazuje počet sekvencí, kterými je každá pozice pokryta. Místa s nedostatečným pokrytím byla doplněna pomocí Sangerova sekvenování.

Tyto oblasti s nízkým pokrytím ($<10\times$) bylo tedy nutno doplnit (jednalo se typicky o geny pro tRNA). Chybějící sekvenční údaje jsem získala Sangerovým sekvenováním obou PCR fragmentů s využitím vybraných primerů, které nasedaly v pokrytých úsecích před oblastmi s malým pokrytím z RNA-Seq.

4.9. Alignment mitogenomů

Získané sekvenční mitogenomů byly ručně zarovnány, zkontrolovány a anotovány podle získané referenční sekvenční normy (KF918859; Bendová et al., 2014) v programu BioEdit 7.0.9.0. Výsledný soubor s alignmentem mitogenomů všech linií mtDNA pak byl použit pro fylogenetické analýzy.

4.10. Amplifikace a klonování sekvencí *cob*

Vzhledem k tomu, že analýza dat z mtDNA nepotvrdila existenci baskické linie (sekvenční *cob* normy z Baskicka získané z kompletního mitogenomu neodpovídaly sekvencím, které byly popsány Deffontaine et al. 2009; viz kapitola Výsledky), rozhodla jsem se ověřit, zda při

použití stejných primerů a PCR programů, které užili autoři článku, nedochází k amplifikaci *numts*, které by odpovídaly sekvencím baskické linie.

Toto ověření bylo provedeno u dvou vybraných jedinců z Baskicka (č. 4 a 5), a také u dvou jedinců z Kalábrie (č. 8 a 9), u kterých byly již dříve pozorovány známky amplifikace nespecifických sekvencí.

4.10.1. Primery

Byly použity dva sety primerů (viz tabulka v příloze č. 2), které pro amplifikaci oblasti *cob* normíka poprvé použili Deffontaine et al. (2005) a Kotlík et al. (2006) a od té doby byly využívány v mnoha pracích, které se zabývají fylogenetikou normíka (Deffontaine et al. 2009, Searle et al. 2009, Wójcik et al. 2010, Boratyński et al. 2011, Colangelo et al. 2012). První amplikon, získaný postupem dle Deffontaine et al. (2005) s primery LCLE1 a MCLE3, má délku 1079 bp a pokrývá nukleotidy 14232-15310 v mtDNA. Druhý amplikon, získaný dle Kotlíka et al. (2006) pomocí primerů CytB24 a Clethr1172, má délku 1170 bp a odpovídá oblasti 14219-15388, při čemž zasahuje až do trnT (viz Bendová et al. 2014).

4.10.2. DNA polymeráza a master mix

Pro PCR byl použit PPP Master Mix (Top-Bio), výsledná koncentrace jednotlivých složek byla následující:

1x PPP Master Mix (75 mM Tris-HCl, pH 8.8, 20 mM $(\text{NH}_4)_2\text{SO}_4$, 0.01% Tween 20, 2.5 mM MgCl_2 , 200 μM dATP, 200 μM dCTP, 200 μM dGTP, 200 μM dTTP, 2.5 u Taq-Purple DNA), 0.3 μM pro forwardní i reverzní primer a ~ 10 ng/ μl DNA v 25 μl PCR směsi.

4.10.3. PCR program

Pro oba sety primerů byl použit stejný program:

Počáteční denaturace - 94°C/ 2 min, denaturační krok - 94°C/ 30 s, annealing – 51°C/ 40 s, extenze – 72°C/ 1 min, cyklus 30× opakován, finální extenze – 72°C/ 10 min.

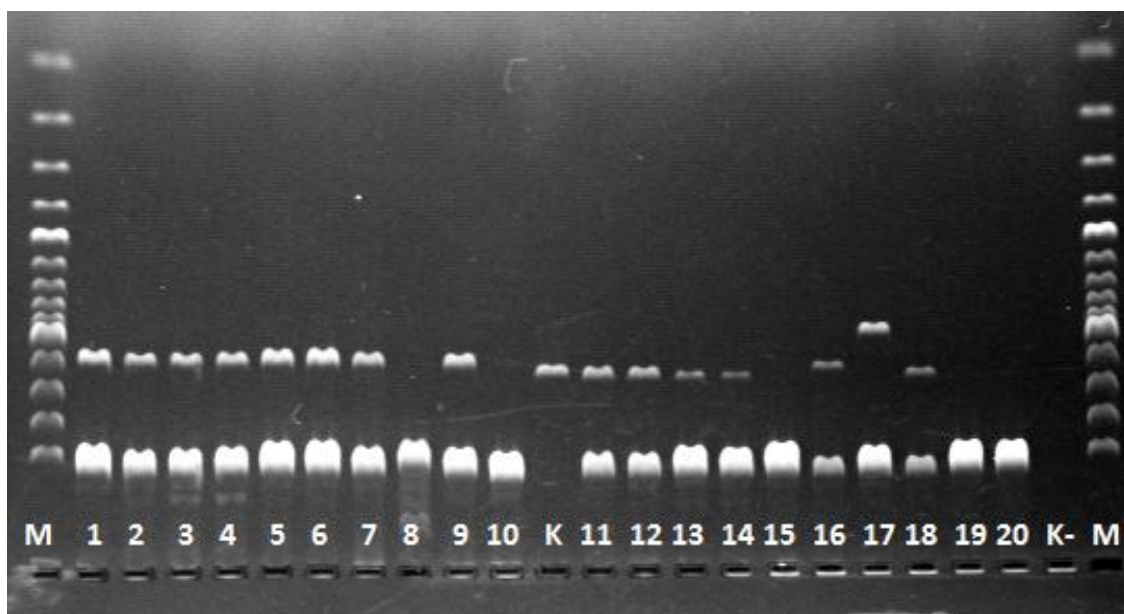
4.10.4. Purifikace a sekvenace

Získané amplikony byly nejprve purifikovány pomocí komerčního kitu (QIAquick PCR Purification Kit, Qiagen) a poté osekvenovány Sangerovou metodou s primery LCLE2 a RCYTB u amplikonů získaných dle Kotlíka et al. (2006) a LCLE1 a MCLE3 u amplikonů získaných dle Deffontaine et al. (2005).

4.10.5. Klonování a sekvenace inzertu

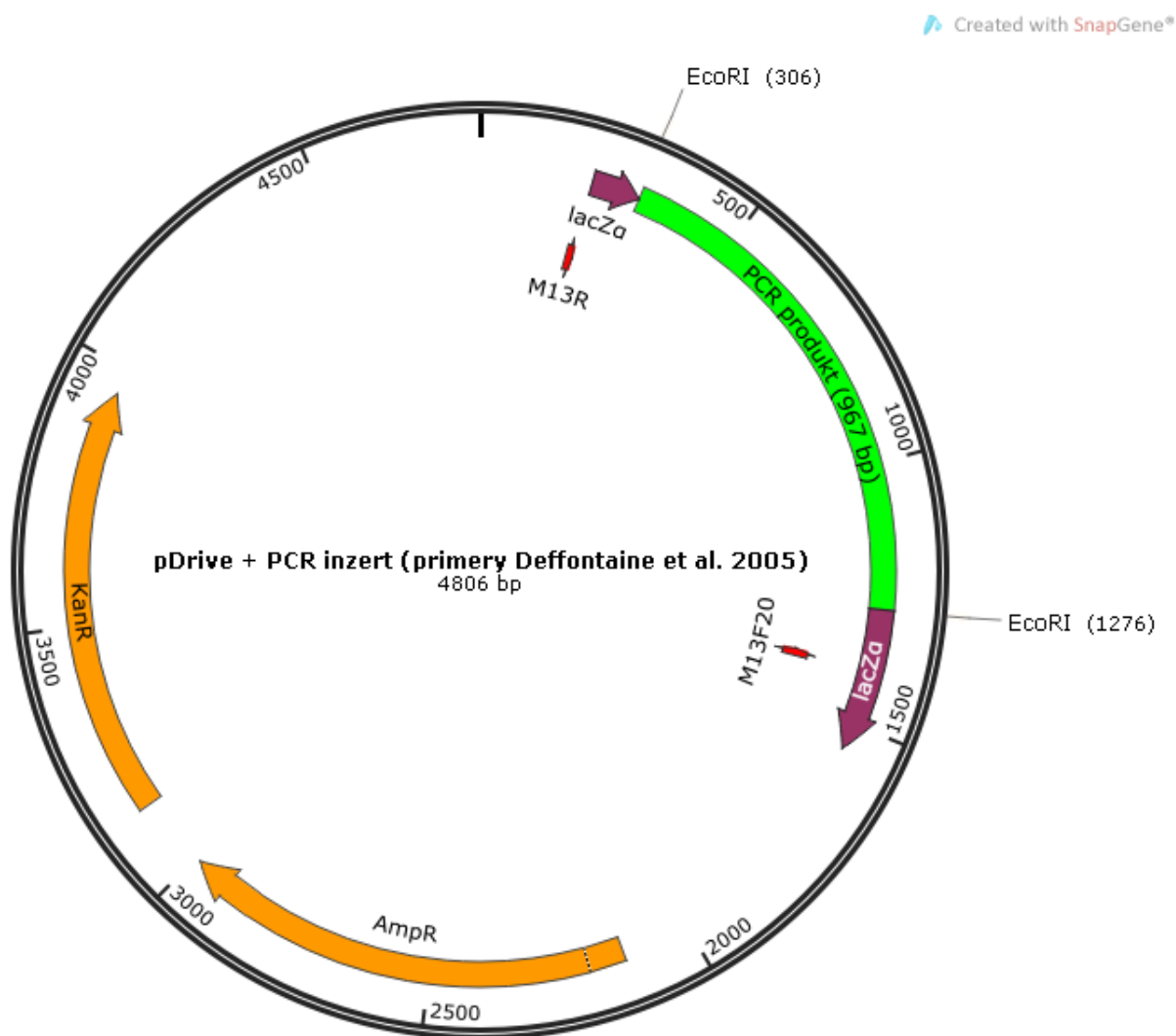
Protože v získaných sekvencích *cob* byly zjištěny vícenásobné polymorfismy (dvojité píky), amplikony byly dále klonovány pomocí PCR Cloning Kitu (Qiagen) s využitím vektoru pDrive v kompetentních buňkách *E. coli lacZΔM15* transformovaných metodou tepelného šoku. Následně byla u bakterií provedena selekce na inzert na médiu s ampicilinem a X-gal a z kolonií s inzertem byla izolována plasmidová DNA pomocí komerčního kitu (QIA Prep® Miniprep Kit, Qiagen).

Ověření přítomnosti a délky vloženého fragmentu v plasmidu bylo provedeno štěpením endonukleázou EcoRI, která štěpí plasmid po stranách inzertu, a následnou analýzou naštěpené DNA na gelu (viz obr. 12).



Obr. 12. Ověření správné délky restrikčních fragmentů po klonování štěpených EcoRI. Jamky 1-7 a 9: zaklonovaný inzert je správné délky (obsahuje PCR fragment získaný primery Deffontaine et al. 2005). Jamky 11-14, 16 a 18: zaklonovaný inzert je správné délky (obsahuje PCR fragment získaný primery Kotlíka et al. 2006). Jamky 8, 10, 15, 19 a 20: Klony jsou bez inzertu. Klon 17 obsahuje neznámý inzert (délka neodpovídá ani jednomu PCR fragmentu). Pozitivní PCR kontrola (PCR fragment získaný primery Kotlíka et al. 2006): K, negativní kontrola: K-.

Plasmidová DNA z klonů, které obsahovaly správně dlouhý fragment (6-9 klonů na každý amplikon), byla následně osekvenována pomocí primerů M13F20 a M13R komplementárních k sekvenci vektoru, které nasedají na obou stranách inzertu (viz obr. 13).



Obr. 13. Zjednodušená mapa plasmidu pDrive po zaklonování PCR fragmentu *cob*. Zobrazena jsou místa nasedání sekvenačních primerů (M13F20 a M13R, červená barva), restrikční místa pro EcoRI, gen pro barevnou selekci na médiu (LacZα; fialově), geny pro selekci na rezistenci vůči antibiotikům (AmpR, KanR; oranžově) a samotný zaklonovaný inzert (zeleně), zde jako příklad amplikon získaný s primery dle Deffontaine et al. (2005). Mapa byla vytvořena v programu SnapGene®.

Získané sekvence byly srovnány se sekvencemi získanými přímým sekvenováním PCR produktů a ručně zkontrolovány, aby se odhalily případné chyby způsobené „rekombinací“ při PCR (Cronn et al. 2002).

4.11. Fylogenetické analýzy

4.11.1. Testy neutrální evoluce

4.11.1.1. TreeSAAP

Název programu je odvozen z anglického *Selection on Amino Acid Properties using phylogenetic Trees*, tedy faktu, že studuje „působení selekce na vlastnosti AMK s využitím fylogenetických stromů“. Testy neutrální evoluce prováděné v tomto počítačovém programu pracují se zarovnanými kódujícími sekvencemi DNA a pomocí matematického modelu MM01 s implementovanými modifikacemi (McClellan et al. 2005) stanovují pravděpodobnostní hodnoty změn široké škály známých fyzikálně-chemických vlastností AMK v těchto sekvencích. Na základě těchto hodnot potom program stanovuje nejen přítomnost pozitivní selekce v daných oblastech kódující sekvence, ale rovněž určuje specifické vlastnosti AMK v těchto místech, což umožňuje zhodnotit evoluci v jednotlivých genech (Woolley et al. 2003).

Jinými slovy, program vyhodnocuje dvě základní proměnné:

- 1) Pravděpodobnost výskytu nesynonymní AMK záměny v daném místě AMK řetězce na základě výpočtu náhodného rozložení nesynonymních AMK záměn ve fylogenii, čili to, jak je pravděpodobné, že k zvýšenému výskytu záměn dojde náhodou.
- 2) Význam vyskytujících se záměn na základě vyhodnocení, jaký má změna v dané pozici v AMK řetězci důsledek pro fyzikálně-chemické vlastnosti tohoto místa. Čím radikálnější je záměna, tím vyšší skóre (na škále od 1 do 8) jí program přidělí. Autoři (Woolley et al. 2002) doporučují věnovat pozornost pouze pozicím, které mají skóre 6-8.

Na základě kombinace výsledků obou proměnných pak program vyhodnocuje, zda se ve fylogenii vyskytuje větev, která by mohla být pod vlivem pozitivní selekce (tj. obsahuje více radikálních záměn AMK, než by se očekávalo).

4.11.1.2. GA-Branch

GA-Branch je jedna z analýz dostupných na serveru Datamonkey (datamonkey.org; viz Delpont et al. 2010). Jejím základem je hledání linií, které vykazují známky selekčního tlaku, pomocí genetického algoritmu (Kosakovsky Pond a Frost 2005a). Zaměřuje se na zodpovězení otázky, kde ve fylogenezi v minulosti došlo k selekci.

Program pracuje tak, že tzv. branch model (tj. model, kde každá větev může mít jiný poměr dN/dS) vybraný genetickou analýzou rozdělí větve na fylogenetickém stromu do tříd podle ω (dN/dS). Větve, které spadají do tříd s $\omega > 1$, jsou pak potenciálně ty, které mohly být v minulosti ovlivněny pozitivní selekcí.

4.11.1.3. SLAC/REL/FEL

Všechny tyto metody jsou součástí analýz dostupných na serveru Datamonkey.org (viz výše). Na rozdíl od GA-Branch však jde o metody, které v alignmentu více sekvencí hledají kodonové pozice, jež by mohly být ovlivněny pozitivní či negativní selekcí (Kosakovsky Pond a Frost 2005b). Jinými slovy, tyto metody slouží k hledání známek pozitivní selekce nikoli na úrovni linií, ale kodonů.

Liší se ve způsobu, jakým stanovují ω pro jednotlivá místa. SLAC (*Single Likelihood Ancestor Counting method*) jednoduše počítá množství synonymních a nesynonymních záměn podél fylogenie, zatímco REL (*Random Effects Likelihood*) předpokládá náhodné rozložení těchto míst podél fylogenie a FEL (*Fixed Effects Likelihood*) počítá pravděpodobnost záměny pro každé místo zvlášť (Kosakovsky Pond a Frost 2005b). Metody tak umožňují odhadnout ω pro každé místo v alignmentu a identifikovat pozice pod vlivem selekce. Autoři doporučují aplikovat na zkoumaný soubor všechny tyto metody pro ověření správnosti výsledků (Kosakovsky Pond a Frost 2005b) – pokud jsou výstupy ze všech tří metod konzistentní, s velkou pravděpodobností lze vyloučit možnost, že se jedná o falešně pozitivní výsledek.

4.11.1.4. PRIME

Analýzy PRIME (*PRoperty Informed Models of Evolution*) podobně jako TreeSAAP pracují s vyhodnocením významnosti AMK záměn pro výsledné vlastnosti daného místa v AMK řetězci. Vycházejí z předpokladu, že radikální nesynonymní záměny AMK (čili takové, které výrazně změni fyzikálně-chemické vlastnosti daného místa v AMK řetězci) jsou poměrně vzácné, zatímco konzervativní záměny (kdy dojde k nahrazení původní AMK jinou, s podobnými vlastnostmi) jsou celkem běžné. Ukazuje se však, že význam a povaha AMK záměn se může lišit také v závislosti na tom, o jaký druh či gen se jedná. Stejná AMK záměna tak může mít různý dopad na vlastnosti daného místa podle toho, v jakém místě se nachází (někdy může být radikální a jindy naopak konzervativní). Je tedy potřeba brát v úvahu kontext, v jakém AMK záměnu vyhodnocujeme (tedy pozici v AMK řetězci). PRIME sestává ze dvou

modelů - Conant-Stadler (Conant et al. 2007) a Atchley et al (Atchley et al. 2005). Oba modely při vyhodnocení radikálnosti záměny berou v úvahu jak změnu fyzikálně-chemických vlastností, tak možný dopad na vlastnosti místa, kterého se tato změna týká (a jejich významnost). Každý k vyhodnocení používá trochu jiný set fyzikálně-chemických vlastností (v obou případech se však jedná o 5 důležitých fyzikálně-chemických vlastností AMK).

4.11.2. Konstrukce fylogenetických stromů

4.11.2.1. Mega 5

Vytvoření prvotních, orientačních fylogenetických stromů metodou Maximum Likelihood (ML⁵) bylo provedeno pomocí programu Mega 6 (Tamura et al. 2013). Pro potřeby tohoto programu byl původní alignment upraven do dvou různých souborů:

1) obsahujícího všechny geny včetně tRNA a rRNA genů.

Všechny geny byly upraveny a otočeny tak, aby měly stejný směr a plnou délku (několik genů v mtDNA se částí své kódující sekvence překrývá a některé geny pro tRNA a gen pro NAD6 mají opačný směr, viz Bendová et al. 2014), a u několika genů byly doplněny stop kodony, které jsou jinak kompletovány při polyadenylaci mRNA (připojením polyA konce dojde ke vzniku kompletního UAA stop kodonu; viz např. Chang a Tong 2012; Bendová et al. 2014).

2) obsahujícího pouze protein-kódující geny bez stop kodonu.

Geny byly seřazeny podle pořadí na mtDNA a upraveny a otočeny tak, aby měly stejný směr a plnou délku.

4.11.2.2. PartitionFinder – hledání nejlepšího schématu evoluce sekvencí

Pomocí programu PartitionFinder (Lanfear et al., 2012) byla vytvořena nejlepší schémata partic. Jako partice jsou v těchto analýzách označeny oddíly či úseky DNA (i nespojitě), které vykazují stejný (resp. statisticky neodlišitelný) evoluční vzor (substituční model).

⁵ Maximum Likelihood je statistická metoda, která vybírá nejvěrohodnější hypotézu vzhledem k datům, tedy takovou, která má nejnižší $-lnL$. V tomto případě je hypotézou fylogenetický strom.

Program vychází z fylogenetického stromu (využila jsem stromu z MEGA 6, ale program jej umí i sám navrhnout), jehož podoba se po celou dobu analýz nemění. To řádově zvyšuje přesnost a umožňuje srovnání mnoha milionů schémat v jedné analýze (Lanfear et al., 2012). Nejprve je pro každý předem definovaný oddíl (první, druhé, třetí pozice kodonu atd.) vybrán nejlepší substituční model, nukleotidové pozice se stejným modelem jsou pak spojeny do tzv. *partic*, a nakonec je spočítána pravděpodobnost každého takového schématu *partic*. Na základě této pravděpodobnosti je schématům přiděleno skóre. Za nejlepší schéma *partic* je pak považováno to, které má nejlepší skóre podle určitého kritéria.

K nalezení schématu s nejlepším skóre se používá tzv. *greedy* algoritmus, což je heuristická metoda, která v prostoru možných hypotéz (zde schémat *partic*) hledá nejlépe skórované hypotézy tak, že ve svém okolí hledá hypotézu s nejlepším skóre, na tu pak přejde a tento postup opakuje, dokud neskončí na hypotéze, kolem níž už žádná hypotéza vyšší skóre nemá (tedy věrohodnostním maximu). Algoritmus se nazývá *greedy* (česky *lakomý*, *hrabivý*) proto, že nikdy při hledání nepřejde na hypotézu s nižším skóre.

Pro tuto analýzu bylo použito nastavení nezávislé délky větví pro různé *partice* (to v praxi znamená, že topologie stromu je identická, ale délky větví jsou počítány nezávisle pro každou z *partic*). Při výběru nejvhodnějších modelů pro jednotlivé *partice* byly použity Akaikeho informační kritérium (AIC), upravené Akaikeho informační kritérium (AICc) a Bayesiánské informační kritérium (BIC).

4.11.3. Analýzy ML se zařazením výsledků z *partic*

4.11.3.1. RAxML

RAxML (*Randomized Axelerated Maximum Likelihood*) je program pro rychlou postupnou i paralelní analýzu umožňující vytvoření i velkých fylogenetických stromů. Využívá heuristických metod k tomu, aby snížil čas potřebný pro nalezení nejlepšího stromu (Stamatakis et al. 2005): vytvoří prvotní strom (pomocí metody parsimonie), potom jej rozdělí na části a optimalizuje podobu těchto částí (nazývaných *subtree*; metoda je tzv. *Lazy Subtree Rearrangement*, podrobně viz Stamatakis a Alachiotis 2010), použije uživatelem vybraný substituční model a k prohledávání prostoru hypotéz používá mechanismus, který umožňuje „krok zpět“, tedy přejítí na hypotézu s nižším skóre (čímž se v konečném důsledku zvyšuje pravděpodobnost, že objeví globální, tedy nejen lokální, maximum věrohodnosti).

Nevýhodou RAxML je, že sice rychle vytvoří strom, ale negarantuje, že takto nalezený strom je správný. Umožňuje však do tvorby stromu zařadit data z jiných analýz – například získané schéma partic. Program však neumožňuje použít různé modely pro různé partic.

Analýza ML mých dat tedy byla provedena v programu RAxML 1.3 (Silvestro a Michalak, 2012) s použitím GTR+G+I modelu pro všechny partic.

4.11.3.2. GARLI

Tento program, který opět vytváří fylogenetické stromy metodou ML, používá „genetický“ algoritmus, v němž se simultánně upravují topologie, délky větví a další parametry. Program tyto parametry upravuje tak, až je výsledkem hypotéza (fylogenie) s maximální věrohodností (Zwickl 2006).

Při analýze mých dat byl použit nejlepší model vybraný pro každou partici a u získaných prvotních stromů byla provedena kombinace optimalizace *subtrees* (menších částí stromu) pomocí algoritmů NNI (*nearest neighbour interchanges*⁶) a SPR (*subtree pruning and regrafting*⁷), přičemž pro každý dataset byla tato analýza provedena několikrát, aby byla ověřena konvergence ke stejné topologii, a každá analýza byla navíc provedena v triplikátech s různou náhodnou počáteční topologií stromu.

4.11.4. Bootstrap analýzy

Bootstrapové analýzy jsou používány ke stanovení spolehlivosti fylogenetických stromů. Je to metoda, která náhodně převzorkovává zkoumaný dataset (sloupce alignmentu, tj. nukleotidové pozice), a hodnotí, zda mohly být výsledky ovlivněny stochastickými efekty.

V případě fylogenetických stromů hodnota bootstrapu (tzv. *bootstrap value*) odpovídá frekvenci, s jakou se daný klád vyskytuje mezi bootstrapovými (tj. randomizovanými) datovými soubory.

⁶ NNI = metoda, která optimalizuje vnitřní topologii stromu tak, že mění pozice dvou sousedních větví a hodnotí, jestli má výsledný strom lepší skóre.

⁷ SPR = česky asi nejlépe „prořezávání a přeroubování jinak“. Metoda, optimalizující vnitřní topologii stromu tak, že v nejlepším stromu oddělí jeden *subtree* (malou část). Tento pak jako celek zkouší přiřadit na všechna možná místa fylogenetického stromu a hodnotí skóre vzniklé topologie.

Ke stanovení podpory větvení v topologii stromů získaných v programech RAxML a GARLI bylo v každém případě použito tisíc bootstrapových vzorků.

4.11.5. Bayesiánské fylogenetické analýzy - MrBayes

Tento typ analýz se snaží odhadnout aposteriorní pravděpodobnost hypotézy (zde fylogeneze) a ne jen její věrohodnost (likelihood) jako je tomu u metod ML. Aposteriorní pravděpodobnost je potom produktem apriorní pravděpodobnosti a věrohodnosti. Spíše než pro schopnost brát v úvahu různé apriorní pravděpodobnosti jednotlivých stromů (které obvykle stejně neznáme a používáme tzv. neinformativní apriorní pravděpodobnosti) je metoda ceněna pro snadné vyhodnocení aposteriorních pravděpodobností. Bayesiánské analýzy využívají algoritmus Markov Chain Monte Carlo (MCMC), jehož cílem je vytvořit vzorek všech hypotéz přicházejících v úvahu, kde se každá hypotéza vyskytuje ve frekvenci odpovídající její aposteriorní pravděpodobnosti. Ve výsledném aposteriorním vzorku fylogenetických stromů pak počty stromů, v nichž se každý klád vyskytuje, odpovídají aposteriorní pravděpodobnosti toho kládu, tedy pravděpodobnosti, že daný klád opravdu existuje.

Jednou z výhod tohoto přístupu je, že výrazně zmenšuje šanci, že naše hypotéza bude představovat pouze lokální maximum – $\ln L$. Hlavní nevýhodou je pak značná výpočetní náročnost a s tím související délka trvání analýzy.

Bayesiánské analýzy byly provedeny v programu MrBayes 3.2.2 (Huelsenbeck and Ronquist, 2001). Každá analýza byla provedena ve dvou paralelních bězích, které probíhaly, dokud nebylo dosaženo zjevné konvergence v MCMC a nebyl shromážděn dostatečně početný aposteriorní vzorek, přičemž každý běh sestával z celkem 4 řetězců MCMC spřažených systémem Metropolis Coupling (algoritmus, který pomáhá dosažení konvergence, podrobněji viz Ronquist et al. 2011)⁸. Délka analýzy byla 5 milionů generací, přičemž každých 100 generací byl aktuální fylogenetický strom zapsán do aposteriorního souboru a *burn-in* (čili část výsledků, které jsou získány v počáteční fázi analýzy před dosažením stacionarity MCMC a jsou vyhozeny) byl 500 stromů.

⁸ v praxi to znamená, že řetězce si občas vymění aktuální strom, aby náhodou neprohledávaly jen oblasti lokálních optim. Celá analýza se pak označuje jako MCMCMC nebo MC³ – *Metropolis Coupled Markov Chain Monte Carlo*.

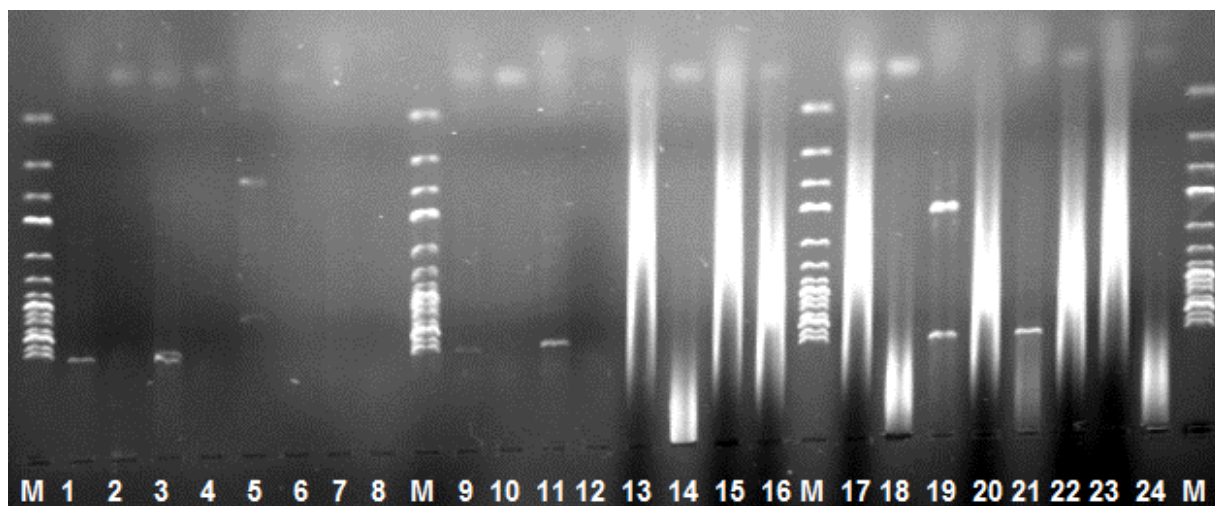
5. Výsledky

5.1. Amplifikace mtDNA pomocí long-range PCR

Amplifikace mtDNA pomocí long-range PCR a následná Sangerova sekvenace se na jednu stranu ukázaly jako poměrně elegantní a rychlý způsob, jak získat kompletní sekvence mitogenomů, avšak ne u všech vzorků byla tato metoda zcela bezproblémová.

V první řadě bylo poměrně obtížné při navrhování primerů pro PCR najít dvě oblasti, které by byly dostatečně konzervativní mezi druhy příbuznými normálnímu člověku, aby se co nejvíce snížilo riziko, že navržené primery nebudou na některé divergentní linie normálního fungovat, a zároveň byly ve vhodné vzdálenosti (tedy takové, aby se vzniklé fragmenty překrývaly minimálně 200 bp na obou koncích a jejich délka byla do 10 kbp).

Pro získání obou fragmentů mitogenomu jsem používala dvě polymerázy – LA DNA polymerázu (Top-Bio) a LongAmp Taq DNA polymerázu (New England Biolabs). Jako lepší se nakonec osvědčila LongAmp Taq DNA, především proto, že na rozdíl od LA DNA poskytovala vyrovnané výsledky bez smearu (nečistot v produktu). První fragment u referenční sekvenace se mi sice podařilo získat pomocí LA DNA polymerázy, ale tato polymeráza se záhy ukázala jako nevhodná. Při kontrole výsledků PCR pomocí elektroforézy se totiž na gelu objevoval smear, a to jak ve vzorcích, tak pozitivních i negativních kontrolách. Tento problém nezmizel ani poté, co jsem vyměnila všechny používané chemikálie za nové, vyzkoušela zcela novou šarži LA DNA polymerázy, nechala vyčistit pipety, použila zcela čerstvé chemikálie při výrobě gelu a rovněž se nechala zkontrolovat při práci kolegy z laboratoře, abych vyloučila možnost kontaminace. Jakmile byla doba extenze při PCR delší než 8 minut, začal se ve výsledcích objevovat smear (viz obr. 14).



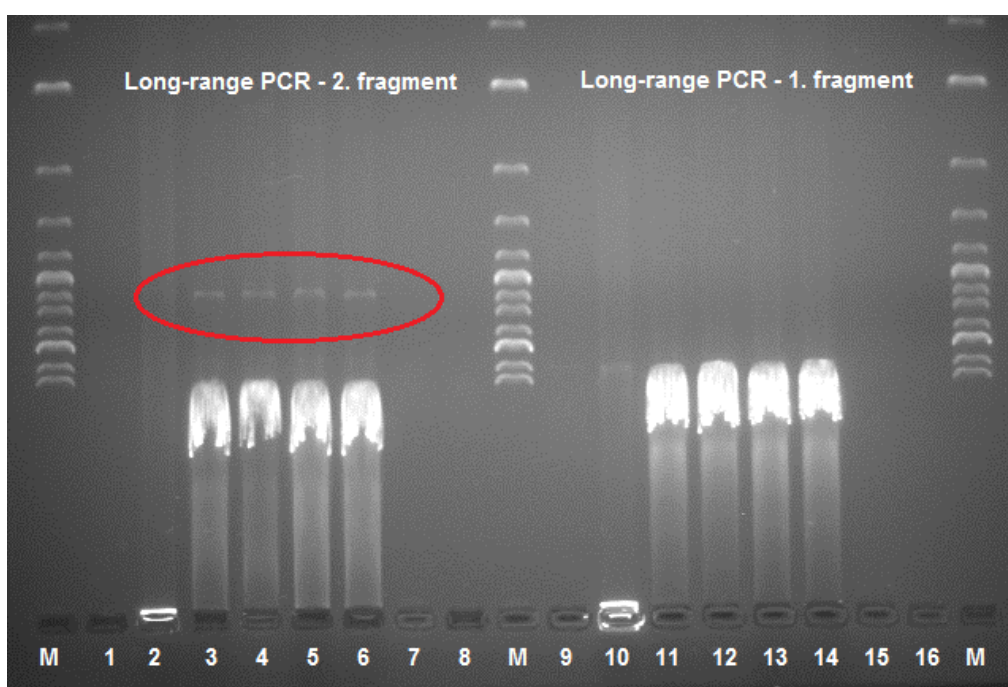
Obr. 14. Výsledky amplifikace LA DNA polymerázou, testování různých párů primerů. Liché číslo je vždy vzorek, sudé pak negativní kontrola. V řadách 1-12 byl amplifikován 1. fragment (extenze 8 min), v řadách 13-24 potom 2. fragment (extenze 10 min 30 s). Je zde jasně vidět, že při delší extenzi se na gelu objevuje smear.

Kvůli tomuto zásadnímu problému (kdybych zkrátila dobu extenze, nestačila by na amplifikaci DNA požadované délky) jsem tedy vyměnila LA DNA polymerázu za LongAmp Taq. Ta již fungovala naprosto spolehlivě, a proto jsem při amplifikaci všech dalších vzorků zůstala u ní.

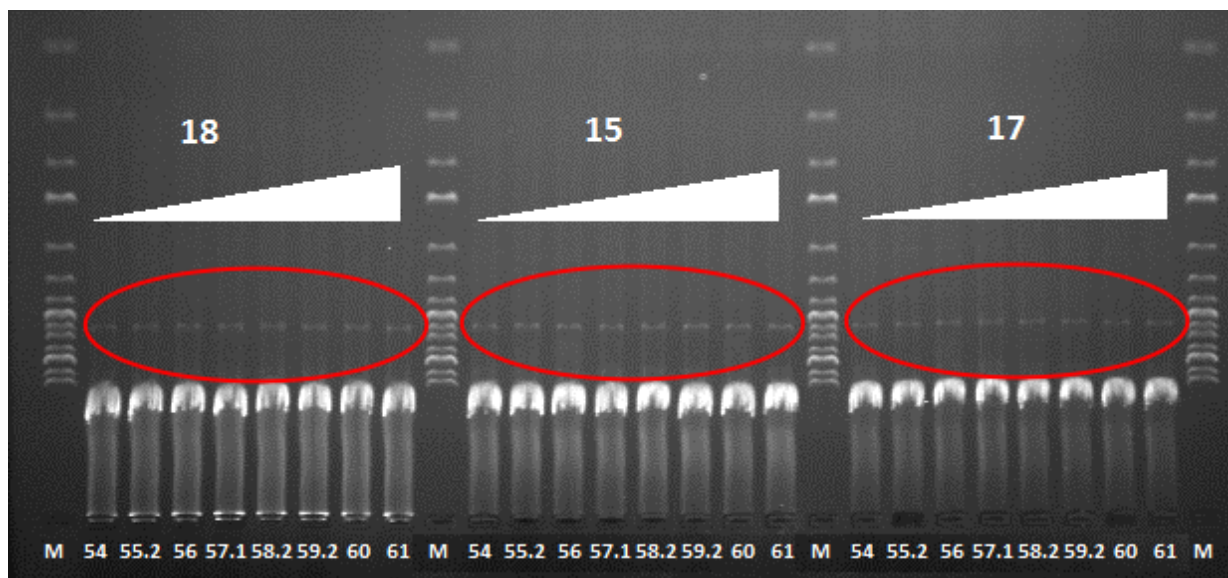
Long-range PCR výborně fungovala na většině vzorků, které byly odebrány a fixovány přímo členy naší laboratoře. Tkáně těchto vzorků byly získány z živochytných pastí a ihned po odběru zakonzervovány v etanolu. Bylo u nich tedy předejito degradaci DNA, což je zásadní krok pro úspěšnou PCR zvláště v případech, kdy se pracuje s dlouhými úseky DNA.

Ostatní vzorky (u nichž bohužel neznám podrobnosti ohledně odběru) už neposkytovaly zdaleka tak vyrovnané výsledky. Někdy ani kvalitní celková DNA, která se po izolaci na gelu jevila jako nedegradovaná, neposkytovala dobré výsledky. Jako možné vysvětlení se nabízí, že ve vzorcích byly přítomny inhibitory PCR. Protože etanol, v němž byly skladovány vzorky, mohl být jedním z těchto inhibitorů, zkoušela jsem u vzorků, kde nebyla amplifikace úspěšná, novou izolaci DNA s prodloužením kroku, v němž dochází k odpařování etanolu ze tkáně. Někdy ani opakovaná izolace nebyla zárukou dostatečně dobrého výsledku PCR, ale celkově vzorky, u nichž byla použita čerstvě izolovaná DNA, dávaly lepší výsledky. U některých vzorků, které nejprve při PCR dávaly horší výsledky, pomohlo, když se PCR program prodloužil o 3 cykly. Abych získala dostatečnou koncentraci DNA k purifikaci a následnému sekvenování, bylo u některých vzorků (např. zástupců španělské linie) nutno použít vícenásobnou PCR amplifikaci a PCR produkty slévat.

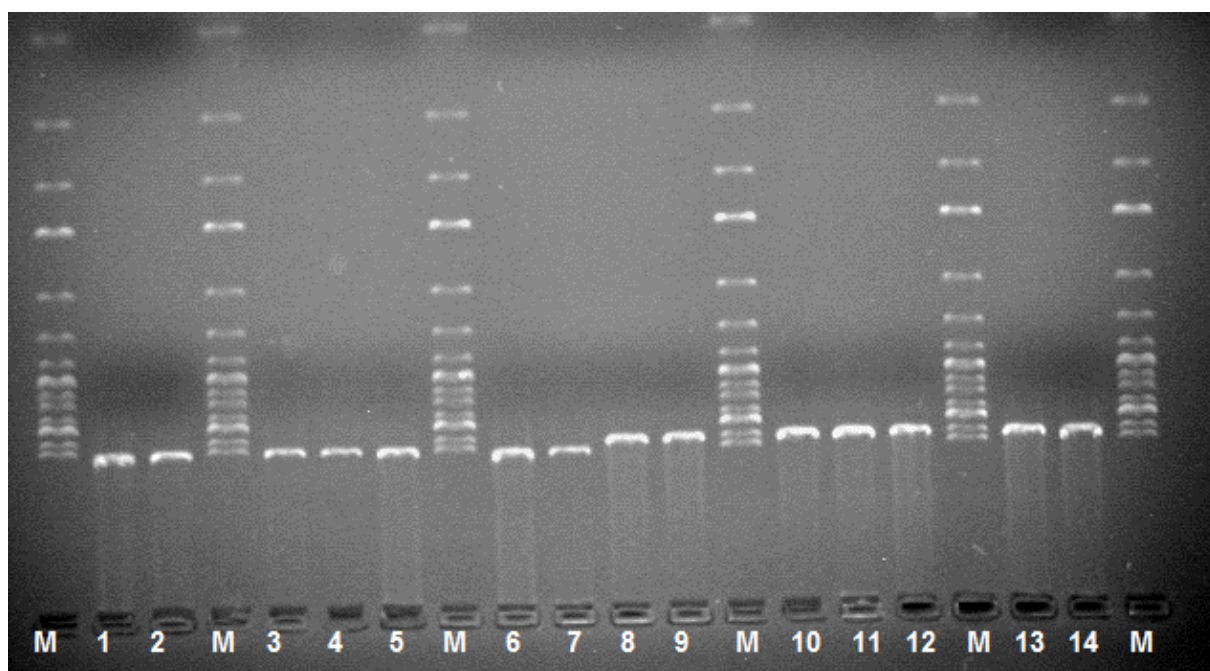
Dalším problémem, který jsem řešila i u vzorků, u nichž docházelo k amplifikaci dostatečného množství produktu, bylo, že při amplifikaci druhého fragmentu vznikl jeden nežádoucí pruh (viz obr. 15). Situaci jsem se snažila řešit pomocí PCR v termálním gradientu (viz obr. 16), ale neúspěšně. Nakonec se však ukázalo, že purifikace, během níž dochází k velkým ztrátám koncentrací DNA, stačila na to, aby koncentrace tohoto nežádoucího produktu klesla natolik, že při analýze na gelu již nebyl patrný; koncentrace cílového produktu přitom zůstala dostatečná (obr. 17). Vzhledem k tomu, že při sekvenování se nevyskytovaly problémy, které by ukazovaly na přítomnost nespecifického produktu (dvojité píky, vysoké pozadí apod.), dále jsem se tímto problémem nezabývala.



Obr. 15. Výsledky PCR amplifikace obou fragmentů mtDNA. 1-2, 7, 9-10, 15 - neúspěšné amplifikace; 3-6 a 11-14 – úspěšná amplifikace; 8 a 16 – negativní PCR kontrola. U 2. fragmentu je zakroužkován nežádoucí pruh.



Obr. 16. Teplotní gradient PCR u 2. fragmentu na třech různých vzorcích (18, 15 a 17), jednotlivé vzorky jsou odděleny markerem; čísla pod jamkami uvádí teplotu annealingu. Jde vidět, že změna teploty nemá vliv na přítomnost nežádoucího pruhu.



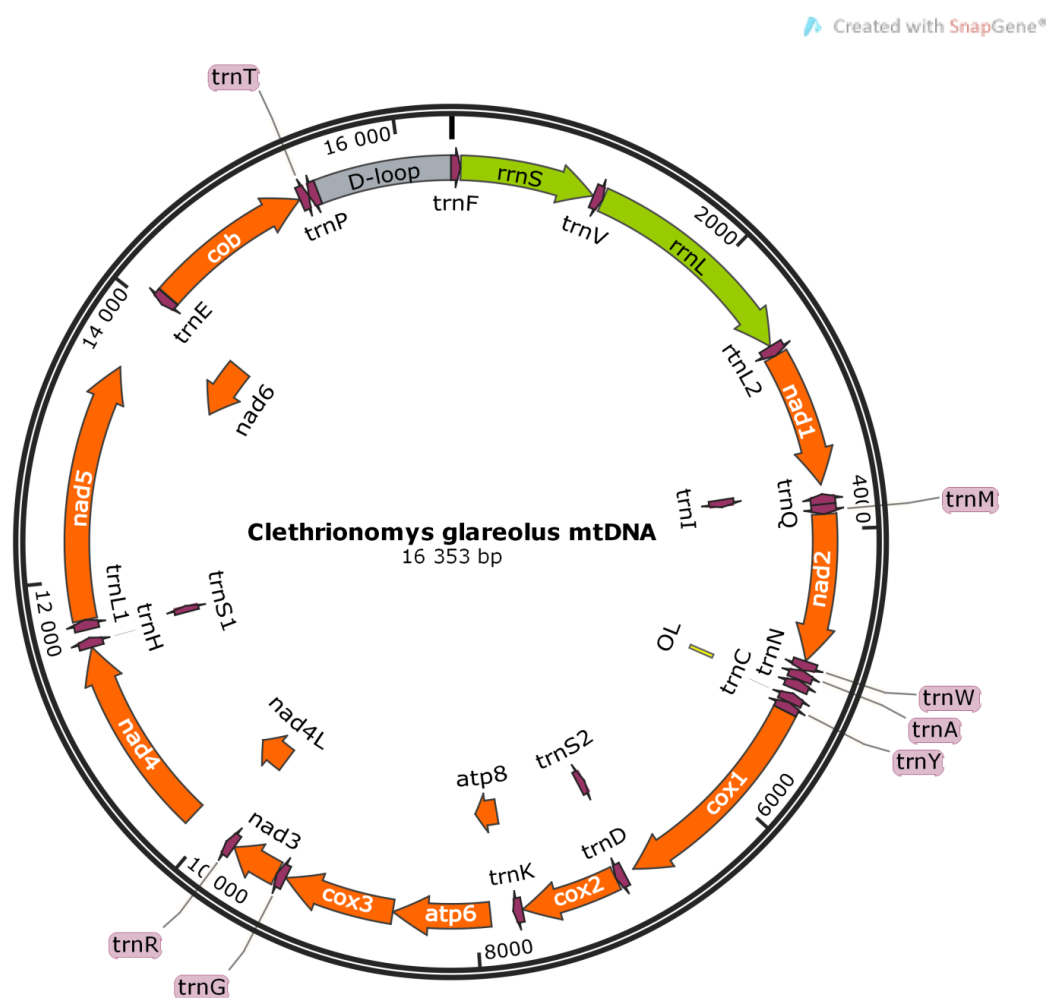
Obr. 17. Purifikované PCR fragmenty. Jamky 1-7 - fragment 2, Jamky 8-14 - fragment 1. Je vidět, že problematický pruh u fragmentu 2 po purifikaci již není vůbec patrný.

Co se týká úspěšnosti amplifikace, u všech jedinců se výrazně lépe amplifikoval první fragment. Přisuzuji to menší délce fragmentu (7800 vs. 9500 bp), neboť amplifikační primery fungovaly i na sekvenci *C. rutilus* a po přesekvencování jsem v oblastech nasedání primerů nezjistila ani u linie, jejichž získávání bylo nejvíce problematické (španělská linie), žádné rozdíly v sekvenci, které by mohly mít vliv na správné nasedání primerů.

5.2. Sekvence mitogenomů

5.2.1. Vytvoření referenční sekvence norníka rudého (primer walking)

Metodou primer walking byla získána první anotovaná sekvence mtDNA norníka rudého (podrobně viz Bendová et al. 2014, příloha P5). Anotace byla vytvořena na základě anotace korejského norníka *C. regulus* (NC_016427, Kim a Park 2012) a myši domácí *M. musculus musculus* (NC_010339). Geny pro tRNA byly anotovány za pomoci programu MITOS (Bernt et al. 2013). K vytvoření anotace a kontrole získané sekvence byla také využita data z RNA-Seq, především pro anotaci genů se stop kodony neúplně kódovanými v DNA, které jsou doplněny až při polyadenylaci (TAA nebo Taa; viz Bendová et al. 2014). Schéma mitochondriální DNA norníka viz obr. 18.



Obr. 18. Mitochondriální DNA norníka rudého (*C. glareolus*). Většina genů je kódována na těžkém vlákne (šípky ve směru hodinových ručiček). Protein kódující geny jsou znázorněny oranžovou, geny pro tRNA fialovou, rRNA geny zelenou, D-Loop (kontrolní oblast) šedou. Vytvořeno v programu SnapGene®.

5.2.2. Získ kompletních mitogenomů všech mtDNA linií (Sangerovo sekvenování)

Celkem bylo získáno 33 sekvencí (včetně první anotované referenční, viz přílohy P1, P5 a P6) kompletních mitogenomů norníka. Výsledné mitogenomy získané Sangerovou metodou byly tvořeny v průměru 31 překrývajícími se sekvencemi. U mitogenomů, které byly získané kombinací Sangerova sekvenování a RNA-Seq byla Sangerovým sekvenováním vyřešena pouze místa, kde RNA-Seq neposkytlo dostatečné pokrytí ($<10\times$ pokrytí); šlo průměrně o 5-6 sekvencí, viz obr. 11.

Získané mitogenomy jsou mírně variabilní ve své délce – toto rozpětí je od 16 351 bp do 16 354 bp. Tato variabilita je dána jednak malými inzercemi či delecemi (tzv. *indely*) v D-Loop, dále se potom někteří jedinci liší v délce rRNA (12S i 16S) a některých tRNA (pro fenylalanin, izoleucin, serin a threonin).

Výchozí alignment všech sekvencí měl délku 16 364 bp. Upravený alignment (v němž byly všechny geny otočeny do stejného směru a jejich překrývající se části byly duplikovány tak, aby každý gen byl kompletní a byly doplněny stop kodony) měl pak délku 16 437 bp. Alignment, který obsahoval kompletní protein-kódující geny bez stop kodonů (otočené do stejného směru), pak měl délku 11 358 bp.

Každý norník měl unikátní sekvenci mitogenomu, až na dvě výjimky, a to norníky č. 6 a 7 z oblasti Kalábrie (Itálie) a 4 a 5 z Baskicka, u nichž byly sekvence identické. Mitogenom norníka rudého z Rany v Norsku náležející k uralské linii sekvencí odpovídal mtDNA norníka tajgového (*C. rutilus*).

Průměrná odlišnost⁹ mezi sekvencemi získaných jedinečných 31 mitogenomů byla 0,018 (přičemž rozpětí zjištěných hodnot bylo 0,003–0,072). Nejvíce odlišná byla sekvence norníka z Rany (0,072 – tedy asi $4\times$ odlišnější, než průměr). Pokud byla tato sekvence ze srovnání vyřazena, průměrná odlišnost mezi sekvencemi klesla na 0,014 (0,003–0,027).

⁹ Odlišnost sekvencí je definována procentem rozdílných bazí v DNA.

5.3. Fylogenetické analýzy

Všemi fylogenetickými analýzami byly získány takřka stejné výsledky, bez ohledu na to, která metoda, algoritmus nebo počítačový program či kritérium výběru partic (AIC, AICc či BIC) byly použity. Topologie fylogenetických stromů získaných z alignmentů kompletních mitogenomů a protein-kódujících genů byly shodné.

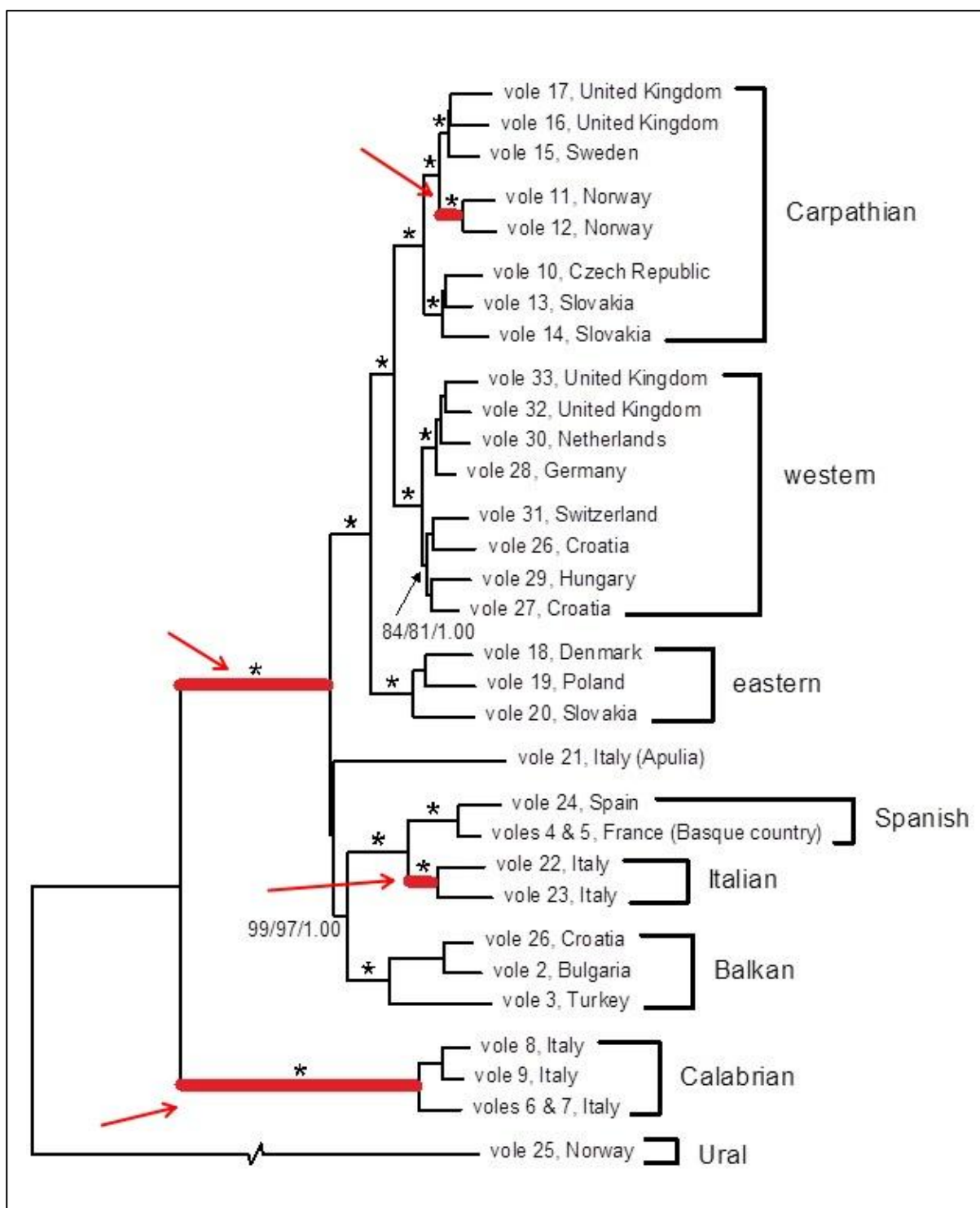
Nejlepší schéma partic pro alignment kompletních mitogenomů obsahovalo podle kritéria BIC soubor 1 zahrnující *rrnS*, *rrnL*, tRNA geny, D-Loop a první kodonové pozice všech protein kódujících genů, druhé kodonové pozice *atp8* a *nd6* a třetí kodonové pozice *nd6* (jako nejlépe odpovídající model zde vyšel HKY+I), potom soubor 2 zahrnující druhé kodonové pozice všech protein kódujících genů s výjimkou *atp8* a *nd6* (zde rovněž vyšel model HKY+I) a soubor 3 zahrnující třetí kodonové pozice všech kódujících genů vyjma *nd6* (model GTR+G).

Mé výsledky jednoznačně potvrdily všechny klády dříve popsané podle sekvencí *cob* – západní, východní, karpatský, italský, španělský, balkánský, Gargano a kalábrijský. Jedinou výjimku představuje baskická mtDNA linie. Genom dvou norníků, které jsme z Baskicka dostali s tím, že nesou mtDNA baskické linie, totiž překvapivě spadl do stejného kládu spolu se španělskou linií. Všechny klády bez výjimky mají maximální statistickou podporu, kterou mohou poskytnout zvolené analýzy – 100% v bootstrapových ML analýzách a 1.00 v Bayesiánské aposteriorní pravděpodobnosti (viz obr. 19).

Fylogenetické stromy byly kořeny pomocí sekvence mitogenomu norníka z lokality Rana, který je označen jako zástupce uralského kládu, a který díky *mitochondrial capture* (Tegelström 1987) nese sekvenci *C. rutilus* (Deffontaine et al. 2005). Kalábrijský klád tvoří sesterský klád k ostatním liniím *C. glareolus*. Tři středomořské klády – španělský, italský a balkánský – společně tvoří statisticky dobře podpořený klád, po němž se postupně odvětvují východní klád a zbývající dva kontinentální klády: západní a karpatský. Toto další větvení je taktéž dobře statisticky podpořeno. Karpatský a západní klád potom společně tvoří klád s vysokou podporou, přičemž oba jsou dále vnitřně rozčleněny. V západním kládu je možné rozlišit dvě skupiny genomů – první z lokalit na východ od Alp a z Alp (Chorvatsko, Maďarsko a Švýcarsko) a druhou z lokalit ležících západně od Alp (Německo, Nizozemsko a Británie). V karpatském kládu pak pozorujeme tři skupiny, opět s vysokou statistickou podporou: jedna tvořená genomy z Karpat na Slovensku a v České Republice, druhá sestávající z genomů z Británie a Švédska, a třetí, v níž nacházíme genomy z Norska (obr. 19). Genom získaný

z Italské Apulie (Gargano) nespadá k žádnému dosud popsanému kládu, ale odvětvuje se z bazální trifurkace spolu se středomořským a kontinentálním kládem.

Velkým překvapením bylo zjištění, že genomy dvou jedinců z Baskicka nevytvořily samostatný klád, jak by se dalo předpokládat na základě předchozích studií používajících jako marker *cob* (Defontaine et al. 2009), ale vždy tvořily jednu linii spolu se vzorky španělské linie. Neočekávané byly tyto výsledky proto, že vzorky pocházely z lokality (Armendarits), v níž byl dosud zaznamenán výskyt pouze norníků z baskické linie (na základě *cob*; Male et al. 2012). Provedla jsem proto další PCR analýzy na *cob*, abych tyto výsledky ověřila, viz níže.



Obr. 19. Fylogenetický strom (ML) z kompletních mitogenomů. Čísla uvedená u větví jsou (postupně) hodnoty z bootstrapových analýz z programů GARLI a RAxML a Bayesiánské aposteriorní pravděpodobnosti. U větví označených hvězdičkou byla získána maximální podpora větvení, čili 100% frekvence v bootstrapových analýzách a aposteriorní pravděpodobnost 1.00. Délka větve definující uralskou linii (*C. rutilus*) byla zkrácena na polovinu pro přehlednější zobrazení výsledků. Ve větvích, které jsou označeny červenou tučnou čarou a šipkou, bylo objeveno zvýšené množství významných nesynonymních záměn. Všechny dříve popsané mtDNA linie mají v této fylogonii vysokou podporu, s výjimkou linie baskické, která klastruje do jednoho kládu společně s linií španělskou, zároveň je jasně vidět, že britští norníci nesou mtDNA jak západní tak karpatské linie.

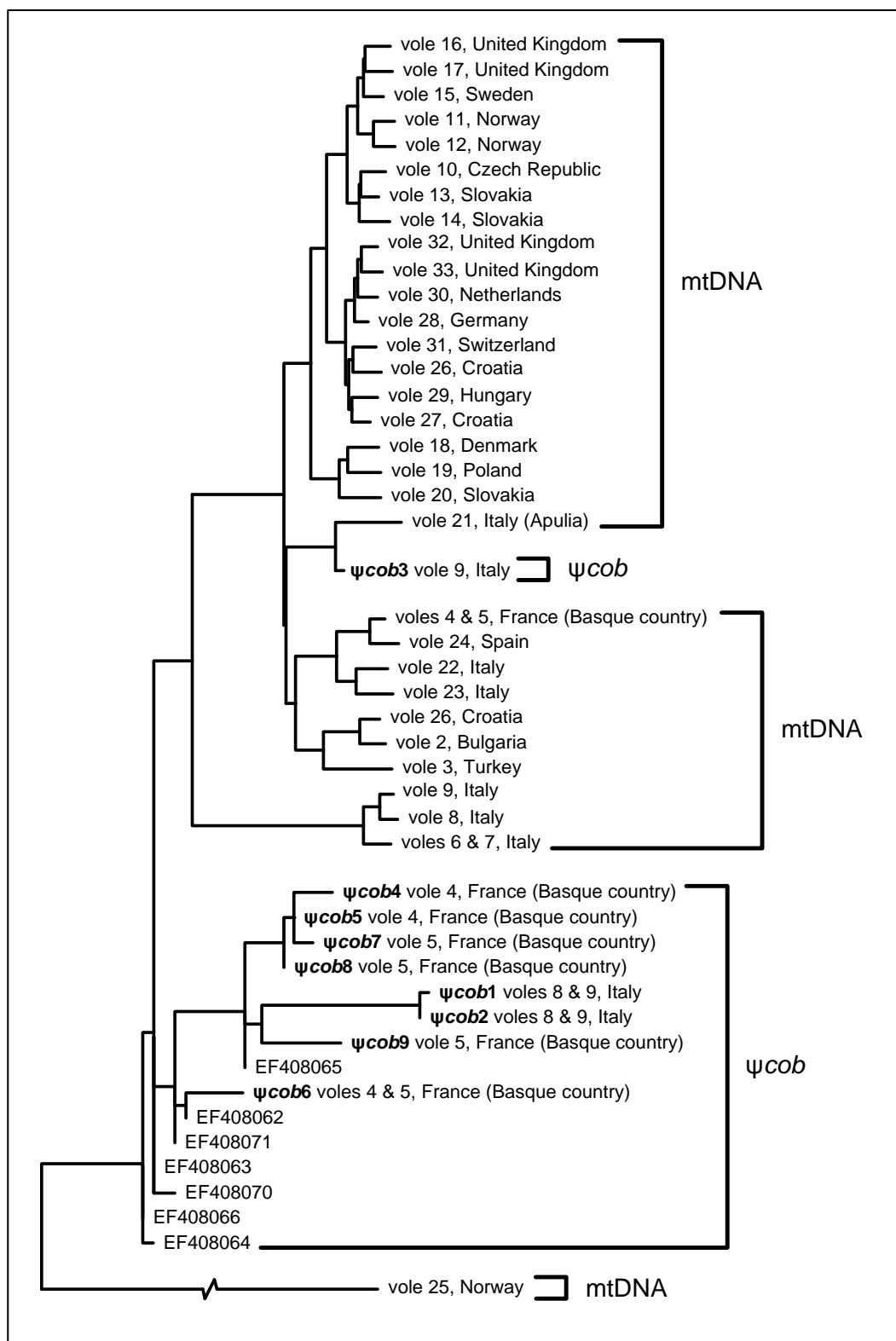
5.4. Analýza sekvencí *cob*

V sekvencích získaných přímým Sangerovým sekvenováním *cob* amplikonů norníků z Baskicka a Kalábrie bylo zjištěno velké množství nejasných míst s překrývajícími se píky. Přitom nehrálo roli, zda byl PCR produkt získán primery z práce Deffontaine et al. (2005) nebo Kotlíka et al. (2006).

U norníků z Baskicka (č. 4 a 5) bylo celkem ze čtyř PCR amplikonů získáno sedm jedinečných sekvencí, které byly každá přítomna v několika klonech získaných z téhož amplikonu. Žádná z těchto sekvencí neměla posunutý čtecí rámec ani neobsahovala předčasný stop kodon. Jedna z takto získaných sekvencí byla nalezena u obou norníků z Baskicka, a byla získána z obou PCR amplikonů (s primery podle Deffontaine et al. 2005 i podle Kotlíka et al. 2006). Srovnáním se sekvencí *cob* z kompletního mitogenomu stejných jedinců bylo zjištěno, že jde o identické sekvence. Tímto se potvrdilo, že oba norníci z Baskicka opravdu patří k španělské linii mtDNA. Z ostatních šesti sekvencí získaných klonováním amplikonů byla jedna společná pro oba norníky (avšak amplifikovala se u každého jiným párem primerů, viz tab. P4 v příloze), dvě (amplifikované pomocí primerů Kotlíka et al. 2006) byly nalezeny pouze u norníka č. 4 a tři pak jen u norníka č. 5 (amplifikované primery Deffontaine et al. 2005).

Amplikony norníků z Kalábrie (č. 8 a 9) také po osekvenování vykazovaly nejasnosti v sekvencích. Stejně jako u baskických vzorků, i zde byla nalezena sekvence, která je shodná pro oba jedince a odpovídá sekvenci *cob* z jejich kompletního mitogenomu. Kromě této sekvence však byly klonováním identifikovány další tři jedinečné sekvence (přičemž všechny se amplifikovaly oběma páry primerů), z nichž dvě byly společné pro oba testované jedince, a jedna byla nalezena pouze u norníka č. 9 (viz tab. P4 v přílohách).

Těchto devět sekvencí (šest pocházejících z norníků z Baskicka a tři z norníků v Itálii), které neodpovídaly sekvencím z *cob*, bylo použito na vytvoření alignmentu spolu se sekvencemi kompletních mitogenomů a se sedmi dříve publikovanými sekvencemi (dostupnými v databázi GenBank pod čísly EF408062-EF408066, EF408070 a EF408071), náležícími dle Deffontaine et al. (2009) k baskické linii. Při fylogenetické analýze ani jedna z těchto devíti sekvencí nespádla do jednoho kládu se španělskou linií (obr. 20) ani s jinou známou linií mtDNA. Místo toho všech devět sekvencí vytvořilo spolu s publikovanými sekvencemi z databáze GenBank samostatnou skupinu u báze fylogenetického stromu (obr. 2).



Obr. 2. Fylogenetický strom obsahující sekvence klonovaných amplikonů. Vzorky jsou popsány dle původu sekvence – u sekvencí pseudogenů popsáných na základě výsledků z klonování amplikonů je uveden znak pro pseudogen ψ , u sekvencí mitogenomů je uvedeno číslo jedince, z něhož sekvence pochází (vole = norník), u sekvencí baskické linie z GenBank je uvedeno databázové číslo. Sekvence pseudogenů vytváří spolu s dříve popsánými sekvencemi z Baskicka společný klastř u báze fylogenetického stromu, zatímco sekvence kompletní mtDNA norníků z Baskicka spadají do společného kládu se španělskou linií.

5.5. Testy neutrální evoluce

Branch model vybraný analýzou GA-Branch rozdělil větve na fylogenetickém stromu z mitogenomů do pěti tříd podle ω (dN/dS), přičemž hodnoty ω pro jednotlivé třídy se pohybovaly v rozmezí od $\omega = 0.00$ do $\omega = 0.30$. Žádná z větví tedy nepřesáhla svým ω hodnotu 1 (což by naznačovalo přítomnost pozitivní selekce v průběhu evoluce mitogenomu), a u všech větví byla zjištěna nulová pravděpodobnost, že mají $\omega > 1$. Na úrovni kodonů vykazovalo $\omega > 1$ pouze jedno místo v alignmentu ze všech 3786 kodonů, přičemž tento výsledek byl podpořen všemi třemi metodami (SLAC, $P < 0.05$; FEL, $P < 0.01$; REL, Bayes faktor 2510.90), což je důležité kritérium mající za cíl eliminovat případné falešně pozitivní výsledky (jak doporučují Kosakovsky Pond a Frost, 2005b).

Toto místo s $\omega > 1$ se nachází v sekvenci proteinu *cob* a jde o záměnu, která vede ke změně z původního alaninového (Ala) kodonu (GCC) na threoninový (Thr) kodon (ACC). V celém stromu bylo objeveno 10 větví, u nichž se vyskytla tato záměna. Analýza PRIME toto místo označila jako místo významně konzervované z hlediska polarizace postranního řetězce (8.1 u Ala ve srovnání s 8.6 u Thr; vlastnost 2 z Conant et al., 2007; $\alpha_2 = 20.00$, $P < 0.001$) a celkového náboje (jak Ala, tak Thr jsou při fyziologickém pH nenabitě; kompozitní vlastnost 5 z Atchley et al., 2005; $\alpha_5 = 15.00$, $P < 0.001$). PRIME nicméně ukazuje na to, že se v tomto místě objevila radikální změna v kompozitní vlastnosti 3 významně častěji (tj. vždy, když došlo k záměně Ala kodonu za Thr kodon), než by odpovídalo neutrální evoluci ($\alpha_3 = -8.16$, $P < 0.001$). Tyto výsledky naznačují, že by toto místo mohlo být předmětem působení pozitivní selekce na některý z atributů této kompozitní vlastnosti. Molekulární objem (např. objem postranního řetězce), který patří mezi atributy s nejvyšším skóre (Atchley et al. 2005), se nabízí jako možný kandidát.

Metoda TreeSAAP pak identifikovala zvýšený výskyt velmi významných (radikálních až extrémně radikálních, stupeň 6-8) změn aminokyselinových vlastností pro větve vedoucí ke čtyřem kládům (obr. 19). Substitute, které měly vliv na redukci povrchu proteinu přístupného rozpouštědlu (vodě) po sbalení proteinu (*Solvent Accessible Reduction Ratio*; stupeň 8; $z = 3.44$, $P < 0.001$) a na rovnovážnou konstantu ionizace COOH skupiny (*Equilibrium Constant (Ionization of COOH)*; stupeň 6; $z = 3.48$, $P < 0.001$), se objevily podél větve definující kalábrijský klád častěji, než by odpovídalo neutralitě. Oproti tomu podél bazální větve definující klád tvořený ostatními evropskými liniemi se významně častěji vyskytovaly substitute ovlivňující afinitu k C-konci (*Power to Be at the C-terminal*; stupeň 6;

$z = 3.31$, $P < 0.001$), viz obr. 19. Substitute ovlivňující energii interakcí parciálních nábojů nebo volných elektronových párů (*Long Range Non-Bonded Energy*, stupeň 6; $z = 3.12$, $P < 0.001$) se vyskytovaly s vyšší frekvencí podél větve italského kládu. V karpatském kládu, respektive podél větve definující subklád sestávající z norských mitogenomů, pak byla zaznamenána vyšší frekvence substitucí, které mají za následek změnu tendence k tvorbě α -helikální struktury (*α -Helical Tendency*, stupeň 6; $z = 3.27$, $P < 0.001$). Výše popsanými radikálními změnami byly ovlivněny geny pro *nad4-6* u kalábrijského kládu, *nad5* u k němu sesterském kládu (tvořeného všemi ostatními liniemi), *nad5* a *cob* u italského kládu a *atp6*, *atp8*, *nad3* a *nad5* u norského (sub)kládu.

6. Diskuse

Mitochondriální DNA je jedním z klíčových markerů užívaných ve fylogenetice. Ačkoli její popularita vzhledem ke zvyšující se dostupnosti celogenomového sekvenování oproti dřívějšímu poněkud klesá, stále představuje velice užitečný nástroj pro analýzu evoluční historie organismů a zasazení těchto poznatků do geografického kontextu – jedním důvodem je možnost navázání na obrovské množství prací, které byly na základě různých genů mtDNA dosud publikovány, druhým pak fakt, že i když použití mtDNA jako markeru má svá úskalí, stále se vyznačuje mnoha vlastnostmi, které jsou z hlediska molekulární analýzy výhodné. S velkým rozvojem sekvenačních metod a především neustále klesající cenou sekvenačních analýz navíc konečně začíná připadat v úvahu i možnost pracovat celkem rutinně se sekvencemi kompletních mitogenomů, což bylo ještě před pár lety téměř nemyslitelné kvůli časové, ale především finanční náročnosti s tím spojené. Doba však pokročila a kompletní mitogenomy tak dnes představují velmi komplexní a poměrně snadno dostupný zdroj dat.

6.1. Technická stránka práce

Zvolený postup – tedy amplifikace mtDNA pomocí long-range PCR ve formě dvou překrývajících se fragmentů – považuji za velmi elegantní, avšak má jistá úskalí.

Prvním zdánlivým omezením je to, že amplifikace byla prováděna z kompletní DNA - ani u dlouhých fragmentů DNA nelze zcela vyloučit možnost amplifikace jaderných pseudogenů. Byly totiž zaznamenány i případ přenosu velmi dlouhých úseků mtDNA (Triant a DeWoody, 2007, Lopez et al. 1994) do jaderného genomu. Právě Lopez et al. (1994) zaznamenali u kočky vložení mtDNA úseku o délce ~7,5 kbp, což je skoro polovina kočičího mitochondriálního genomu. Ačkoli tedy amplifikace mtDNA pomocí long-range PCR značně snižuje pravděpodobnost amplifikace jaderných pseudogenů, čistě teoreticky by pro dosažení maximální spolehlivosti metody bylo nejlepší nejprve izolovat mtDNA.

Nezdá se však, že by byly výsledky poznamenány současnou amplifikací pseudogenů. Sekvence získané Sangerovým sekvenováním byly většinou velice kvalitní, a to dokonce i v případě druhého fragmentu, kde se při PCR amplifikoval jeden další úsek DNA (nejspíše proto, že byl přítomen v tak nízké koncentraci, že při purifikaci PCR produktů došlo k jeho kvantitativní eliminaci). Případy, kdy jsem nedosáhla kvalitní sekvence ani opakovaným sekvenováním, byly výjimečné a zpravidla bylo příčinou to, že sekvenační primer nasedal před

oblastí s repetitivní sekvencí, která pak často způsobila posun signálu o jednu či více bazí, a sekvence pak nebyla pro program, který vyvolával báze, dostatečně čitelná.

Na základě výsledků long-range PCR tedy vyvozují, že k dosažení dobrých výsledků není potřeba izolace mtDNA, ale stačí důsledná kontrola získaných sekvencí.

Pokud pracujeme se sekvencemi získanými Sangerovou metodou, je velmi vhodné vždy ručně zkontrolovat, zda sekvence skutečně odpovídá získanému chromatogramu – u problematických sekvencí totiž jenom pouhé vyvolání bazí (*base calling*) nemusí nutně odhalit problém v sekvenci (který se většinou projevuje přítomností míst se smíšeným signálem ze dvou nebo více bazí nebo vysokým sekvenačním pozadím). U konsenzuálních sekvencí získaných NGS je pak ideální pracovat pouze s oblastmi s dostatečným pokrytím (tzn. více než 10, nebo ještě lépe 50 sekvencemi). Nedodržení tohoto kritéria může způsobit, že místo skutečných sekvencí ve výsledku analyzujeme sekvence, které jsou artefakty vzniklými na základě chyb NGS, a výsledky analýz jsou potom zavádějící. Stejně tak je dobré věnovat pozornost i dříve publikovaným sekvencím (např. získaných z databáze GenBank), které používáme pro analýzy, zvláště takovým, které obsahují nevyřešená místa (kódovaná jako „N“).

Jak již jsem zmínila ve výsledcích, v mé práci se potvrdilo, že k amplifikaci dlouhých úseků DNA pomocí PCR je potřeba velmi kvalitní, nenalámaná DNA. MtDNA je sice díky své délce a tvaru ve srovnání s jadernou DNA proti degradaci odolnější, ale přesto se vyplatí postupovat tak, aby se minimalizovalo riziko jejího polámání, tedy omezit pipetování a veškeré manipulace s roztoky DNA na minimum. Nemohu zcela zodpovědně posoudit, jaký způsob konzervace tkání po odběru je pro zabránění degradace DNA nejlepší, protože informace o způsobu odběru mám jen u vzorků, které sbírali mí kolegové z laboratoře. Minimálně však mohu říct, že právě s těmito vzorky, k jejichž získání byly použity živochytné pasti a jejichž tkáň byly ihned po odběru zakonzervovány v etanolu, jsem dosahovala nejlepších výsledků při long-range PCR.

6.2.Mitogenomická fylogeografie norníka rudého v Evropě

Tato práce představuje fylogenetickou analýzu 33 sekvencí kompletních mitogenomů norníka rudého, které byly získány kombinací metod NGS (Illumina) a Sangerovy metody sekvenování. Kompletní mitogenomy prokázaly skvělou schopnost rozlišit nejen různé linie mtDNA, ale také

jasně stanovit jejich vzájemné příbuzenské vztahy. Gen *cob* se ukazuje jako dobrý marker pro primární identifikaci linií, ale co se týká jejich evoluce, neposkytuje již dostatečně silný signál.

Většina fylogenií dosud vytvořených pomocí *cob* norníka měla v lepším případě střední podporu v bootstrapových analýzách; co se týká stanovení vzájemné příbuznosti linií, výsledky byly velmi nejednoznačné. Nejčastěji se vyskytující problém byl ten, že některé linie vždy nevytvářely monofyletické skupiny – tyto komplikace se vyskytly například u západní linie v Deffontaine et al. (2009), španělského kládu v Colagnello et al. (2012), nebo např. u skupiny tří mediteránních kládů (španělského, italského a balkánského), které dohromady neseskupila analýza v Deffontaine et al. (2009).

Proti tomu všechny fylogenetické analýzy, které jsem realizovala na kompletních mitogenomech, poskytly výbornou statistickou podporu pro linie mtDNA norníka, které byly dosud popsány v Evropě, a ukazují jasně jejich vzájemné příbuzenské vztahy – s jednou výjimkou, a to linií baskickou. Ačkoli v práci Deffontaine et al. (2009) se tato linie mtDNA při fylogenetických analýzách jevila jako bazální ke všem ostatním dosud popsaným liniím, při mých analýzách kompletních mitogenomů genomy norníků z Baskicka spadly do jednoho kládu se vzorkem reprezentujícím španělskou linii. Možné příčiny těchto výsledků diskutuji níže.

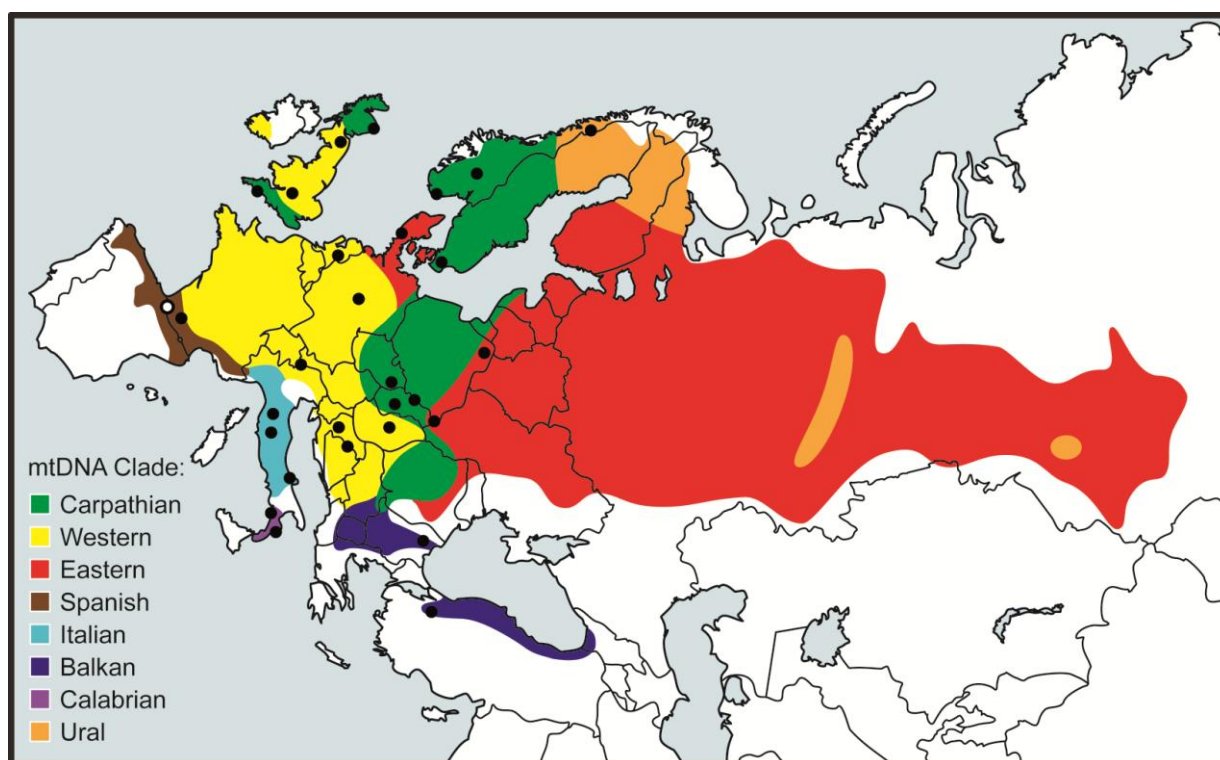
Z mé práce je možno odvodit několik zásadních závěrů pro fylogeografii norníka.

V první řadě je z fylogenetického stromu jasně patrné, že nejstarší linií norníka v Evropě je linie kalábrijská. Tato linie totiž vytváří sesterský klád ke kládu, který zahrnuje všechny ostatní dosud popsané linie mtDNA norníka. Tento fakt spolu s tím, že tato linie nebyla dosud zaznamenána jinde než v Kalábrii, ukazuje na to, že ačkoli svou geografickou polohou měla Kalábrie i v LGM vhodné klima pro přežívání temperátních druhů a sloužila jako glaciální refugium, zřejmě nebyla zdrojem, z něhož by po LGM probíhala postglaciální kolonizace Evropy.

Neméně zajímavý je pak samostatný klád, který vytvořila sekvence norníka z Apulie (podle poloostrova, na němž se nachází, jsem tuto linii nazvala Gargano). Jde rovněž o velice starý klád, sesterský ke všem ostatním dosud popsaným středomořským kládům.

Tři hlavní středomořské klády (španělský, italský a balkánský) dohromady vytváří společný klád, sesterský klád k tomuto je pak tvořen klády západním, karpatským a východním.

Podobný vzor byl již u norníka dříve pozorován na základě krátkého úseku ze sekvence *cob* (Bilton et al. 1998) a je jedním z argumentů proti univerzální platnosti modelu kolonizace Evropy po LGM z jižních refugií. Mé výsledky tuto hypotézu podporují - všechny mediteránní klády se totiž značně liší od ostatních linií, které nacházíme jinde v Evropě (viz obr. 21). Jako nejpravděpodobnější příčinu minimálního příspěvku mediteránních linií k postglaciální kolonizaci severních částí Evropy vidím především geografické bariéry, které mohly velice efektivně bránit rozšíření těchto linií – Pyreneje, Alpy a hory na Balkánském poloostrově. Jižní refugia tak spíše byla útočištěm endemických linií než zdrojem postglaciální kolonizace (Bilton et al. 1998).



Obr. 21. Mapa distribuce mtDNA linií norníka rudého v Evropě (upraveno podle Wójcika et al. 2010). V mapě je nově vyznačena distribuce karpatské linie ve Velké Británii, Norsku a Švédsku, dále také východní linie v Dánsku a západní linie v Nizozemsku. Lokality, původně náležící baskické linii, byly přiřazeny španělské linii (hnědá). Byla přidána rovněž uralská linie, kt. odpovídá sekvenci mtDNA *C. rutilus* (oranžová).

Jako velmi pravděpodobná se naopak na základě mých výsledků jeví postglaciální kolonizace Evropy z refugií ležících mimo Středomoří. Vzhledem ke geografické distribuci linií vyskytujících se v severněji položených lokalitách Evropy se domnívám, že tato refugia byla tři.

První se zřejmě nacházelo v oblasti na severozápadním cípu balkánského poloostrova - odsud se pak dále rozšířila západní linie (jak navrhli Deffontaine et al. 2005). Dnes je tato linie široce rozšířena především v západní Evropě, ale setkáme se s ní také v Británii (viz níže).

Druhé refugium, z něhož pochází norníci náležící východní linii, bylo nejspíše mnohem dále na východě. Jeho přesnější lokalizace je stále otázkou, ale Deffontaine et al. (2005) navrhuje oblast Uralu.

Poslední refugium se pak zřejmě nacházelo v oblasti Karpat, jak navrhuji Kotlík et al. (2006). Z něj pocházející karpatská linie měla zjevně značný podíl na kolonizaci severní Evropy a Británie – mé výsledky jako první ukazují, že norníci z Norska, Švédska a severozápadu Británie nesou sekvence nejvíce příbuzné této linii. Ve fylogenetických analýzách karpatská linie vytváří jasný klád, který se dále dělí na dvě skupiny, z nichž jedna je tvořena mitogenomy norníků ze Slovenska a České Republiky, a druhá pak mitogenomy britských, švédských a norských norníků. Jako vysvětlení fylogeneze karpatské linie se nabízí existence více než jednoho refugia – výše popsané klády by mohly odrážet kolonizaci severní a západní Evropy ze dvou refugií, jednoho položeného severněji a západněji (z něhož vzešla „severní“ větev, zahrnující švédskou, norskou a britskou skupinu) a druhého více na jihu v Karpatech (ze kterého pochází dnešní česko-slovenská větev norníků).

„Severská“ větev karpatské linie se dále dělí na další skupiny, přičemž jako první se od společného předka oddělila skupina norská. K oddělení britských od švédských haplotypů zjevně došlo až později. Tento fylogeografický vzor by se dal vysvětlit tak, že ke kolonizaci došlo ve dvou vlnách (vzhledem ke vzájemné podobnosti linií se domnívám, že z jednoho refugia). V první vlně by bylo kolonizováno Norsko, později pak ve druhé vlně Británie a Švédsko, což by mělo za následek, že mezi předky švédských a britských norníků mnohem déle docházelo ke genovému toku a jsou si tedy ve výsledku příbuznější.

6.3.Baskická linie – pouze artefakt?

Analýzy kompletních mitogenomů potvrdily existenci všech mtDNA linií s výjimkou baskické linie. Oba jedinci z Baskicka, jejichž mitogenomy se mi podařilo získat, sdíleli stejnou sekvenci mtDNA, která odpovídala dříve popsané španělské linii. Bylo by však ukvapené na základě jedné sekvence mtDNA (resp. dvou identických) spekulovat o (ne)existenci baskické linie. Deffontaine et al. (2009) totiž na území Baskicka objevili i španělskou linii a ačkoli v lokalitě,

z níž pocházely mé vzorky, nebyla zaznamenána přítomnost jedinců španělské linie mtDNA (Male et al. 2012), nebyl by výskyt jedinců k ní náležícím v této oblasti nijak zvlášť překvapivý. Rozhodla jsem se proto využít možnosti srovnat sekvence *cob* z kompletních mitogenomů se sekvencemi získanými amplifikací *cob* a zjistit, zda takto nezískám další vodítka, která by mi mohla pomoci vyřešit otázku (ne)existence baskické linie.

Původní práce Deffontaine et al. (2009), v níž byla baskická linie poprvé popsána, je z hlediska fylogenetiky norníka zajímavá, ale současně sporná – došlo zde totiž k pochybení při označení vzorků a tato chyba se bohužel odrazila i v publikovaných závěrech. Na základě této chyby byla například neoprávněně zpochybněna existence karpatské linie (Wójcik et al. 2010). Podpora větvení finálního fylogenetického stromu v této studii byla také poměrně překvapivá – prakticky jediné linie s opravdu výbornou podporou v této studii byly baskická a uralská (která odpovídá sekvenci *C. rutilus* a je tedy spíše sekvencí outgroupovou), zbývající linie pak měly podporu výrazně nižší. Nižší podpora z bootstrapových analýz prováděných na vnitrodruhových fylogenetických stromech vytvořených na základě jednoho genu však není nijak výjimečným jevem, proto se zaměřím spíše na samotnou topologii, čili tvar stromu.

Ve své práci Deffontaine et al. (2009) popsali baskickou linii jako sesterský klád ke všem ostatním dosud popsaným liniím, který se nacházel u báze fylogenetického stromu (kalábrijská linie v té době ještě nebyla popsána). Bazální pozice baskické linie však může být interpretována různě.

Jako jedno možné vysvětlení se nabízí to, které použili autoři: Jedná se o starobylou linii mtDNA, která se oddělila jako první.

Druhé možné vysvětlení, které budu diskutovat níže, je to, že ve skutečnosti vůbec nejde o linii mtDNA, ale o jaderné mitochondriální pseudogeny, tedy jaderné sekvence DNA, které mají určitou sekvenční podobnost s *cob* a při PCR se amplifikují spolu se sekvencemi *cob* z mtDNA nebo dokonce namísto nich.

Jaderné mitochondriální pseudogeny, jak již bylo zmíněno výše, jsou výsledkem přenosu mitochondriálních genů do jádra. Příčina tohoto přenosu stále není zcela vyjasněna (viz výše v přehledu literatury), ale jev samotný byl popsán u mnoha savců (Bensasson et al. 2001). Protože přenesené geny jsou v jaderné DNA podstatně lépe evolučně „zakonzervovány“ (mtDNA má výrazně vyšší mutační rychlost než jaderná DNA), pseudogeny se jeví více

podobné sekvencím mtDNA z doby, kdy došlo k přenosu do jádra, než současné mitochondriální sekvence. Pokud tedy vytvoříme fylogenetický strom ze sekvencí pseudogenů a sekvencí, které odpovídají skutečným mitochondriálním genům, sekvence pseudogenů se mohou stahovat k bázi stromu.

V čeledi hrabošovitých byl výskyt pseudogenů podrobně popsán např. u hraboše východoevropského *Microtus rossiaemeridionalis* (Triant a DeWoody 2007). V chromatogramech sekvencí *cob* norníků z Baskicka a Itálie, které jsem získala amplifikací za využití primerů z předchozích fylogenetických studií (Kotlík et al. 2006, Deffontaine et al. 2009), však bylo mnoho míst, které se nepovedlo uspokojivě vyřešit ani opravným přesekvenováním. Možnost, že by šlo o heteroplasmické jedince, sice byla zvážena, ale vzhledem k tomu, že jde o poměrně vzácný jev a u získaných PCR produktů byla přítomnost dvojitého píku častější, než by odpovídalo přítomnosti mutačně příbuzných haplotypů v případě heteroplasmie, jednoznačnou volbou bylo rozřešit otázku těchto nejasných míst za pomoci klonování.

Kromě sekvencí, které odpovídaly sekvenci *cob* získané z kompletního mitogenomu, byly z každého zkoumaného jedince izolovány dvě až čtyři různé sekvence pseudogenů, což ukazuje na to, že přenos genů do jádra byl u některých jedinců spojen také s duplikací pseudogenu (počet sekvencí vylučoval, že se jedná o alelický polymorfismus jednoho lokusu). Post-inzerční duplikace jaderných pseudogenů byly pozorovány rovněž u hraboše východoevropského (Triant a DeWoody, 2007) a zdá se, že ani u jiných druhů, konkrétně kočky domácí (Lopez et al. 1994), nejsou nijak výjimečnou záležitostí.

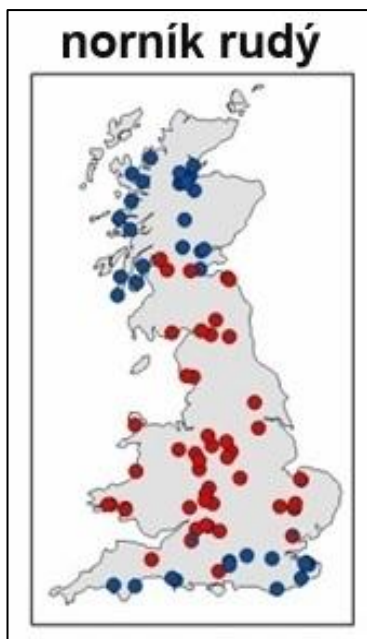
Na fylogenetickém stromu, vytvořeném ze získaných sekvencí pseudogenů, sekvencí kompletních mitogenomů zástupců všech linií a vybraných sekvencí jedinců z GenBank, kteří byli Deffontaine et al. (2009) popsáni jako náležící k baskické linii, jsou jasně patrné známky přitahování získaných sekvencí pseudogenů spolu se sekvencemi jedinců z Baskicka k bázi stromu (viz obr. 20). Z toho, že sekvence pseudogenů klastrují u báze fylogenetického stromu a vyskytují se u více linií (minimálně španělské a kalábrijské), lze usuzovat na to, že k přenosu části mtDNA do jaderného genomu muselo dojít již před rozrůzněním všech linií a lze tedy očekávat, že pseudogeny budou přítomné u všech dnešních linií. Sekvence pseudogenu, která klastruje do jednoho kládu spolu s jedincem z linie Gargano, ukazuje též na to, že v průběhu evoluce došlo k přenosu části mtDNA do jádra minimálně dvakrát. Možnou příčinou, proč k amplifikaci pseudogenů při PCR dochází ve větší míře jen u některých linií, může být mutace

v místech, kde nasedají amplifikační primery, nebo například různá míra post-inzerční tandemové amplifikace sekvencí pseudogenů v jaderné DNA (a tedy různé množství pseudogenové DNA dostupné pro PCR).

Při důkladnější kontrole sekvencí z databáze GenBank, které dle Deffontaine et al. (2009) náleží k baskické linii, jsem u pěti ze sedmi sekvencí zaznamenala minimálně jedno místo s nevyřešenou bazí („N“), což ukazuje na problémy se čtením signálu při sekvenaci (*base calling*) a rovněž podporuje myšlenku, že kromě *cob* v případě těchto vzorků mohlo dojít k amplifikaci jiných sekvencí.

Na základě těchto výsledků a výsledků z analýz kompletních mitogenomů se tedy domnívám, že popsaná baskická linie není skutečnou linií, ale sekvence v databázi GenBank přiřazené této linii v práci Deffontaine et al. (2009) představují soubor pseudogenů či smíšených sekvencí mtDNA a pseudogenů. Nejlepší by pochopitelně bylo osekvenovat kompletní mitogenomy jedinců, jejichž sekvence *cob* byly publikovány v GenBank a označeny za baskickou linii, ale tyto vzorky již bohužel nebyly k dispozici (vzorky, s nimiž jsem pracovala, byly získány od autorů původní publikace a pocházely ze stejné lokality, nicméně šlo o jiné jedince).

6.4.Mitochondriální linie norníků v Británii



Obr. 22. Rozšíření dvou mtDNA linií (data z *cob*; dle Searle et al. 2009). Karpatská linie (potomci prvních kolonistů) modře, západní linie (potomci druhých kolonistů) červeně. Data z kompletních mitogenomů vykazují stejný vzor.

Data z kompletních mitogenomů potvrdila výskyt dvou různých linií norníků v Británii, přičemž jejich geografická distribuce odpovídá „Keltskému lemu“ pozorovanému na základě *cob* (Searle et al. 2009, srv. obr. 21 a 22). Tyto dvě linie, karpatská a západní, představují potomky populací, které kolonizovaly Británii v době mezi ústupem pevninského ledovce na konci LGM (cca 19 000 AC) a zatopením pevninského mostu, který spojoval Británii se zbytkem Evropy (přibližně 8 000 AC). Jak jsem již popsala v přehledu literatury, z dnešního rozšíření těchto dvou linií lze usuzovat na to, že kolonizace Británie probíhala ve dvou vlnách (viz Searle et al. 2009).

Jako první do Británie přišli norníci ze severních refugií. Mé výsledky jasně ukazují, že tyto první kolonisté jsou potomky karpatské mitochondriální linie, jejíž výskyt byl dosud

zaznamenán pouze na evropském kontinentu. Tato data tedy představují významnou podporu pro hypotézu severních refugií temperátních druhů v Evropě jako zdroje kolonizace.

V druhé vlně pak byla Británie kolonizována norníky západní linie. Fylogenetické stromy podporují myšlenku, že pocházejí ze stejného glaciálního refugia jako populace ve střední Evropě, která náleží k této linii. Zmíněné refugium se nejspíše nacházelo více na jihu než refugium, z něhož pochází karpatská linie norníků v Británii; Deffontaine et al. (2005) a Kotlík et al. (2006) jako možnou oblast jeho umístění navrhuji severozápad Balkánského poloostrova či předhůří Alp. Tito druzí kolonisté ze západní linie však z nějakého důvodu začali původní populace vytlačovat, takže dnes nacházíme norníky z karpatské linie mtDNA na severním a jižním okraji ostrova, zatímco západní linie se vyskytuje především v centrální oblasti ostrova.

Vzhledem k tomu, že populace norníků karpatské linie pochází ze severních refugií, dá se předpokládat, že byly lépe adaptovány na drsné klimatické podmínky. Částečně tuto hypotézu podporuje i fakt, že norníci z karpatské linie se nacházejí spíše v centrální a severní Evropě, kde je klima podstatně studenější, než na západě Evropy. Mitochondrie by přitom v této adaptaci mohly hrát významnou roli. Nacházejí se totiž ve zvýšeném množství v hnědé tukové tkáni, která hraje významnou roli v adaptaci savců na chladné klima (podrobněji viz např. review Cannon a Nedergaard, 2004). Protonový gradient, který mitochondrie vytvářejí, je totiž v hnědé tukové tkáni savců přeměňován přímo na teplo; to pak může být využito k ohřívání organismu, aniž by docházelo k svalovému třesu (což je další možný způsob, jakým se savci brání podchlazení). Spojitost mezi haplotypem mtDNA a právě schopností vytvářet teplo netřesovou termogenezí byla pozorována u bělozubky tmavé (Fontanillas et al. 2005); zaměřila jsem se tedy při testech adaptivní evoluce na protein kódující geny mtDNA, které se účastní dýchacího řetězce a vytváření protonového gradientu v mitochondriích.

Žádné známky adaptivní evoluce v mtDNA jsem však ani u karpatské, ani u západní linie vyskytující se v Británii ve svých testech nenašla. Pokud tedy při kolonizaci ostrova hrály adaptivní rozdíly roli (a domnívám se, že je to velmi pravděpodobné), patrně se nevyskytovaly na úrovni genů kódovaných mtDNA. Z jaderných genů se jako velmi pravděpodobný kandidát, který mohl zvyšovat konkurenceschopnost druhých kolonistů, nabízí hemoglobin, zmíněný již dříve v přehledu literatury (viz Kotlík et al. 2014).

Hemoglobin byl totiž jedním z genů, na nichž Searle et al. (2009) popsali Keltský lem, a nedávno publikovaná práce mých kolegů (Kotlík et al. 2014) ukazuje, že v tomto genu se u norníků v Británii vyskytují dvě alelické varianty, lišící se záměnou AMK serinu za cystein v hlavním β -řetězci. Tato záměna má funkční dopad v podobě zvýšené odolnosti červených krvinek proti oxidativnímu stresu, a mohla by tak poskytovat selekční výhodu při kolonizaci. Tuto alelickou variantu nese západní linie, čili potomci druhých kolonistů. Zdá se tedy, že selekční výhodu má právě tato linie a Keltský lem, který pozorujeme v Británii, je spíše odrazem toho, jak daleko zatím pokročila kolonizace ostrova západní linií, než známkou adaptace populací, které momentálně nacházíme v severních částech Británie, na horší podmínky.

Samozřejmě, že hemoglobin je pouze jeden kamínek v rozsáhlé mozaice fyziologických procesů v buňkách organismu, takže je dost dobře možné, že ačkoli v tomto genu, jak se zdá, má výhodu západní linie norníka, v jiných genech může být za určitých podmínek lépe přizpůsobena zase karpatská linie. Rovněž je potřeba vzít v úvahu, že dnešní klima v Británii je značně odlišné od toho, které zde panovalo v době, kdy probíhala první vlna kolonizace. Z dnešní geografické distribuce linií sice můžeme usuzovat na to, jak tento proces probíhal, ale zároveň je třeba mít na paměti, že vidíme pouze konečný stav, tedy linie, které v dané lokalitě přetrvaly do současnosti, a jejich rozšíření je především odrazem adaptace na současné podmínky. Proto by bylo vhodné u linií norníka na tomto území prozkoumat, jaké se u nich vyskytují alelické varianty dalších kandidátních genů pro adaptaci (např. cytochromu c a jiných mitochondriálních proteinů kódovaných v jaderné DNA, případně dalších nemitochondriálních genů).

6.5. Testy adaptivní evoluce

Ve fylogenetickém stromu nebyla objevena žádná větev, která by měla $\omega > 1$, z čehož lze usuzovat na to, že evoluce žádné linie mtDNA nebyla pod vlivem pozitivní selekce. Bylo však objeveno jedno místo v *cob*, 22. kodon, jež mělo $\omega > 1$, a analýza PRIME ukazuje na to, že zde dochází k významné změně molekulárního objemu, a tato změna se objevuje příliš často na to, aby se její výskyt dal vysvětlit jen jako náhodný. Podstatou této změny je záměna GCC kodonu (alanin) na kodon ACC (threonin), a objevila se celkem desetkrát v celé fylogonii, minimálně jednou v prakticky každé linii a u outgroupu (sekvenci odpovídající uralskému kládu). Tato záměna může být známkou toho, že *cob* je pod vlivem pozitivní selekce (viz např. Foote et al. 2011), ale vzhledem k tomu, že se záměna vyskytuje u prakticky všech linií, není příliš

pravděpodobné, že by nějakým způsobem přispěla ke zvýhodnění určitých linií při kolonizaci Evropy. Jak již bylo zmíněno výše, stejně tak výsledky žádného testu nenaznačují, že by distribuce mtDNA linií v Británii byla důsledkem selekční výhody, minimálně co se mitochondriálních genů týká – ani jedné z linií nebylo objeveno ani vyšší množství nesynonymních záměn, ani zvýšené množství radikálních záměn v AMK.

Analýzou sekvencí v programu TreeSAAP však byly identifikovány čtyři větve, podél nichž došlo v průběhu evoluce často k záměnám, které mají velký význam pro fyzikálně-chemické vlastnosti AMK. Nedá se říct, že by pro tyto záměny bylo společné nějaké univerzální schéma, co se pozměněných vlastností týká; v každé větvi byly identifikovány jiné záměny mající vliv na vlastnosti různých genů. Jedna věc však stojí za povšimnutí, a tou je, že větve či klády, v nichž došlo k těmto záměnám, se vyskytují na okrajích areálu rozšíření norníka rudého – v Itálii (včetně Kalábrie) a dále pak v Norsku. Je tedy možné, že tyto signifikantní záměny odráží adaptaci norníka na podmínky, které jsou pro tento druh extrémní (viz podobný příklad u člověka popsany v práci Mishmar et al. 2003).

Významné záměny byly objeveny také ve větvi, která se odděluje od bazální linie spolu s kalábrijským kládem a definuje společného předka všech dnešních mtDNA linií ve střední Evropě. Tyto záměny ukazují na to, že po oddělení této kalábrijské linie prodělal společný předek dnešních zbylých evropských linií poměrně intenzivní evoluci v protein-kódujících genech. Příspěvek těchto záměn k adaptivní evoluci a dalšímu rozšiřování norníka po Evropě však není zcela jasný.

Celkově se tedy dá říct, že u některých kládů norníka rudého se vyskytují významné záměny, které mohly nějakým způsobem přispět k adaptaci tohoto druhu na klimatické podmínky. Pro poněkud méně obecné vyhodnocení, jaký by tyto záměny mohly mít fyziologický význam, by však bylo vhodné u vybraných linií provést hlubší prozkoumání významných záměn pomocí počítačového modelování proteinů, v nichž se záměny nachází, a odhad dopadu těchto změn na schopnost interakce s významnými vazebnými partnery. Stejně tak by bylo vhodné ověřit experimenty *in vivo*, zda je mezi jedinci z obou okrajů areálu rozšíření nějaký zásadní rozdíl ve fyziologických vlastnostech, který by mohl být způsobený mitochondriálními proteiny.

7. Shrnutí

Má práce představuje fylogenetickou analýzu 33 sekvencí kompletních mitogenomů norníka rudého. Data z kompletních mitogenomů prokázaly výbornou schopnost jasně vyřešit fylogenetické vztahy uvnitř tohoto druhu a potvrdily všechny linie popsané v Evropě na základě *cob* s výjimkou baskické, která nejspíše představuje sekvence do této chvíle nerozpoznaných jaderných pseudogenů. Klonováním sekvencí *cob* získaných dvěma běžně užívanými sety primerů bylo zjištěno, že kromě sekvence odpovídající *cob* z mtDNA se těmito primery mohou amplifikovat také sekvence jaderných pseudogenů. Počet rozpoznaných sekvencí a jejich lokalizace na fylogenetickém stromu ukazují na to, že k přenosu genů z mtDNA do jádra došlo v historii minimálně dvakrát a tento přenos byl nejméně v jednom případě následován tandemovými duplikacemi přenesené sekvence.

Bylo identifikováno jedno místo v *cob*, v němž se objevily častější radikální záměny; toto místo však není specifické pro konkrétní linii či větev. Dále byl objeven nadbytek radikálních nesynonymních záměn podél větví, které definují kalábrijský klád, italský klád a norský klád. Ačkoli se u každého kládu tyto záměny týkaly různých vlastností AMK a různých genů, jedno přeci jen mají společné – objevují se u linií, které se vyskytují na okraji areálu rozšíření norníka rudého. Je tedy možné, že vyšší míra významných nesynonymních záměn u těchto kládů souvisí s klimatem těchto stanovišť, z hlediska optimálního prostředí norníka již dost extrémním. U linií, které měly v postglaciální kolonizaci Evropy největší význam, tedy karpatské, západní a východní, však geny mtDNA nevykazují žádné známky funkční adaptace.

Mé výsledky jasně potvrdily, že Británie byla v minulosti kolonizována minimálně dvěma liniemi norníka – karpatskou a západní. První kolonisté pocházeli z karpatského refugia, západní linie kolonizovala Británii až ve druhé vlně asi o 2000 let později. Při porovnání těchto linií jsem nenašla žádné známky funkčních adaptací v mtDNA, které by mohly být příčinou jejich dnešní distribuce (tzv. Keltského lemu); tato příčina tedy leží mimo mtDNA (a pravděpodobně aspoň částečně spočívá ve funkčním rozdílu v genu pro hemoglobin).

Mitochondriální DNA je jako molekulární marker pro fylogenetické studie v mnoha ohledech stále nenahraditelná. S rozvojem sekvenačních technologií a NGS je navíc možné poměrně snadno získat sekvence celých mitogenomů a tedy využít kompletní mtDNA, která jako celek může poskytnout mnohem komplexnější obrázek o vztazích mezi organismy a jejich evoluci, než dosud užívané vybrané úseky.

8. Seznam použité literatury

- Adams, K., & Palmer, J. (2003). Evolution of mitochondrial gene content: Gene loss and transfer to the nucleus. *Mol. Phylogenet. Evol.*, 29(3), 380-395.
- Alberts, B. (2002). The Mitochondrion. In *Molecular biology of the cell* (4th ed.). New York: Garland Science. Online at <http://www.ncbi.nlm.nih.gov/books/NBK26894/>
- Anderson, S., Bankier, A., Barrell, B., De Bruijn, M., Coulson, A., Drouin, J., ... Young, I. (1981). Sequence And Organization Of The Human Mitochondrial Genome. *Nature*, 290, 457-465.
- Anderson, S., Bruijn, M., Coulson, A., Eperon, I., Sanger, F., & Young, I. (1982). Complete sequence of bovine mitochondrial DNA conserved features of the mammalian mitochondrial genome. *J. Mol. Biol.*, 683-717.
- Andersson, S., Zomorodipour, A., Andersson, J., Sicheritz-Pontén, T., Alsmark, U., Podowski, R., ... Kurland, C. (1998). The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature*, 396, 133-140.
- Antunes, A., Pontius, J., Ramos, M., O'brien, S., & Johnson, W. (2007). Mitochondrial Introgressions into the Nuclear Genome of the Domestic Cat. *J. Hered.*, 98(5), 414-420.
- Atchley, W., Zhao, J., Fernandes, A., & Drüke, T. (2005). Solving the protein sequence metric problem. *Proc. Natl. Acad. Sci. U. S. A.*, 102(18), 6395-6400.
- Avise, J. (1987). Intraspecific Phylogeography: The Mitochondrial-DNA Bridge Between Population Genetics And Systematics. *Annu. Rev. Ecol. Syst.*, 18(1), 489-522.
- Avise, J. (2004). Speciation and Hybridization. In *Molecular markers, natural history and evolution* (2nd ed., pp. 374-375). Sunderland, MA: Sinauer Associates.
- Bazin, E., Glémin, S., & Galtier, N. (2006). Population Size Does Not Influence Mitochondrial Genetic Diversity in Animals. *Science*, 312(5773), 570-572.
- Benda, C. (1898). Ueber die Spermatogenese der Vertebraten und höherer Evertebraten, II. Theil: Die Histiogenese der Spermien. *Arch. Anat. Physiol.*, 73, 393-398.
- Bendová, K., Marková, S., Searle, J., & Kotlík, P. (2014). The complete mitochondrial genome of the bank vole *Clethrionomys glareolus* (Rodentia: Arvicolinae). *Mitochondrial DNA* [Epub ahead of print].

- Bensasson, D., Zhang, D., Hartl, D., & Hewitt, G. (2001). Mitochondrial pseudogenes: Evolution's misplaced witnesses. *Trends Ecol. Evol.*, 16(6), 314-321.
- Berg, O., & Kurland, C. (2000). Why Mitochondrial Genes are Most Often Found in Nuclei. *Mol. Biol. Evol.*, 17(6), 951-961.
- Bergstrom, C., & Pritchard, J. (1998). Germline Bottlenecks and the Evolutionary Maintenance of Mitochondrial Genomes. *Genetics*, 149(4), 2135-2146.
- Berlin, S., Tomaras, D., & Charlesworth, B. (2007). Low mitochondrial variability in birds may indicate Hill–Robertson effects on the W chromosome. *Heredity*, 99(4), 389-396.
- Bilton, D., Mirol, P., Mascheretti, S., Fredga, K., Zima, J., & Searle, J. (1998). Mediterranean Europe as an area of endemism for small mammals rather than a source for northwards postglacial colonization. *Proc. R. Soc. B*, 265(1402), 1219-1226.
- Blanchard, J., & Schmidt, G. (1995). Pervasive migration of organellar DNA to the nucleus in plants. *J. Mol. Evol.*, 41(4), 397-406.
- Boore, J. (2006). Requirements and Standards for Organelle Genome Databases. *OMICS: J. Integr. Biol.*, 10(2), 119-126.
- Boratyński, Z., Alves, P., Berto, S., Koskela, E., Mappes, T., & Melo-Ferreira, J. (2011). Introgression of mitochondrial DNA among *Myodes* voles: Consequences for energetics? *BMC Evol. Biol.*, 11(1), 355-355.
- Brown, W., George, M., & Wilson, A. (1979). Rapid Evolution Of Animal Mitochondrial DNA. *Proc. Natl. Acad. Sci. U. S. A.*, 76(4), 1967-1971.
- Cannon, B., & Nedergaard, J. (2004). Brown Adipose Tissue: Function And Physiological Significance. *Physiol. Rev.*, 84(1), 277-359.
- Cedergren, R., Gray, M., Abel, Y., & Sankoff, D. (1988). The evolutionary relationships among known life forms. *J. Mol. Evol.*, 28(1-2), 98-112.
- Chance, B., & Williams, G. (1956). Respiratory enzymes in oxidative phosphorylation. VI. The effects of adenosine diphosphate on azide-treated mitochondria. *J. Biol. Chem.*, 221(1), 477-489.
- Chang, J., & Tong, L. (2012). Mitochondrial poly(A) polymerase and polyadenylation. *BBA - Gene Regulatory Mechanisms*, 1819(9-10), 992-997.

- Claude, A. (1940). Particulate Components Of Normal And Tumor Cells. *Science*, 91(2351), 71-78.
- Claude, A. (1944). The Constitution Of Mitochondria And Microsomes, And The Distribution Of Nucleic Acid In The Cytoplasm Of a Leukemic Cell. *J. Exp. Med.*, 80(1), 19-29.
- Claude, A. (1946). Fractionation Of Mammalian Liver Cells By Differential Centrifugation: I. Problems, Methods, And Preparation Of Extract. *J. Exp. Med.*, 84(1), 51-59.
- Claude, A. (1948). Studies on cells: Morphology, chemical constitution and distribution of biochemical functions. *Harvey Lectures*, 43, 121–164.-121–164.
- Cooper, G. (2000). The Mechanism of Oxidative Phosphorylation. In *The cell: a molecular approach* (2nd ed., p. (online)). Sunderland, MA: Sinauster Associates.
- Cronn, R., Cedroni, M., Haselkorn, T., Grover, C., & Wendel, J. (2002). PCR-mediated recombination in amplification products derived from polyploid cotton. *TAG Theoretical and Applied Genetics*, 104(2-3), 482-489.
- Deffontaine, V., Fontaine, M., Ledevin, R., Renaud, S., Dumont, B., Libois, R., ... Michaux, J. (n.d.). Back to the glacial refugium concept with the bank vole lineage (*Myodes glareolus*) from Gargano, Italia. *Mol. Biol. Evol.*, -(-), Nепубликованá data.
- Deffontaine, V., Libois, R., Kotlík, P., Sommer, R., Nieberding, C., Paradis, E., ... Michaux, J. (2005). Beyond the Mediterranean peninsulas: Evidence of central European glacial refugia for a temperate forest mammal species, the bank vole (*Clethrionomys glareolus*). *Mol. Ecol.*, 14(6), 1727-1739.
- Delport, W., Poon, A., Frost, S., & Pond, S. (2010). Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics*, 26(19), 2455-2457.
- Doolittle, W. (1998). You are what you eat: a gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. *Trends Genet.*, 14(8), 307-311.
- Driggers, W., LeDoux, S., & Wilson, G. (1993). Repair of Oxidative Damage within the Mitochondrial DNA of RINr 38 Cells. *J. Biol. Chem.*, 268(29), 22042-22045.
- Druzhyna, N., Wilson, G., & Ledoux, S. (2008). Mitochondrial DNA repair in aging

and disease. *Mech. Ageing Dev.*, 129(7-8), 383-390.

- Filipi, K., Marková, S., Searle, J., & Kotlík, P. (2015). Mitogenomic phylogenetics of the bank vole *Clethrionomys glareolus*, a model system for studying end-glacial colonization of Europe. *Mol Phylogenet Evol*, 82, 245-257.
- Fontanillas, P., Dépraz, A., Giorgi, M., & Perrin, N. (2005). Nonshivering thermogenesis capacity associated to mitochondrial DNA haplotypes and gender in the greater white-toothed shrew, *Crocidura russula*. *Mol. Ecol.*, 14(2), 661-670.
- Foote, A., Reid, R., Rasmussen, M., Pérez-Gil, M., Vilstrup, J., Stephanis, R., ... Gilbert, M. (2011). Genetic differentiation among North Atlantic killer whale populations. *Mol. Ecol.*, 20(3), 629-641.
- Galtier, N., Nabholz, B., Glémin, S., & Hurst, G. (2009a). Mitochondrial DNA As a Marker Of Molecular Diversity: a Reappraisal. *Mol. Ecol.*, 18, 4541-4550.
- Gissi, C., Iannelli, F., & Pesole, G. (2008). Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species. *Heredity*, 101(4), 301-320.
- Good, J., Hird, S., Reid, N., Demboski, J., Steppan, S., Martin-Nims, T., & Sullivan, J. (2008). Ancient hybridization and mitochondrial capture between two species of chipmunks. *Mol. Ecol.*, 17(5), 1313-1327.
- Gray, M., Sankoff, D., & Cedergren, R. (1984). On the evolutionary descent of organisms and organelles: a global phylogeny based on a highly conserved structural core in small subunit ribosomal RNA. *Nucleic Acids Res.*, 12(14), 5837-5852.
- Gray, M., Burger, G., & Lang, B. (1999). Mitochondrial Evolution. *Science*, 283(5407), 1476-1481.
- Gray, M. (2012). Mitochondrial Evolution. *Cold Spring Harbor Perspect. Biol.*, 4(9), A011403-A011403.
- Gray, M. (2014). The Pre-Endosymbiont Hypothesis: a New Perspective on the Origin and Evolution of Mitochondria. *Cold Spring Harb Perspect Biol.*, 6(3), A016097-A016097. Retrieved October 8, 2014, from <http://cshperspectives.cshlp.org/content/6/3/a016097.full.pdf.html>
- Harijith, A., Ebenezer, D., & Natarajan, V. (2014). Reactive oxygen species at the

crossroads of inflammasome and inflammation. *Front. Physiol.*, 5(352). Retrieved October 27, 2014, from <http://journal.frontiersin.org/journal/10.3389/fphys.2014.00352/full>

- Hewitt, G. (2004). Genetic consequences of climatic oscillations in the Quaternary. *Philos. Trans. R. Soc., B*, 359(1442), 183-195.
- Hogeboom, G., Schneider, W., & Pallade, G. (1948). Cytochemical studies of mammalian tissues; isolation of intact mitochondria from rat liver; some biochemical properties of mitochondria and submicroscopic particulate material. *J Biol Chem*, 172(2), 935-48.
- Holt, I., & Reyes, A. (2012). Human Mitochondrial DNA Replication. *Cold Spring Harbor Perspect. Biol.*, 4(12), A012971-A012971.
- Horáček, I. (2000). Glacial Cycles and Mammalian Biodiversity of Central Europe: Large Scale Migrations or Vicariance Dynamics? *Geolines*, 11, 103-107.
- Huelsenbeck, J., & Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, 17(8), 754-755.
- Johns, G., & Avise, J. (1998). A Comparative Summary of Genetic Distances in the Vertebrates from the Mitochondrial Cytochrome b Gene. *Mol. Biol. Evol.*, 15(11), 1481–1490.
- Kasamatsu, H., & Vinograd, J. (1973). Unidirectionality of Replication in Mouse Mitochondrial DNA. *Nature, New Biol.*, 241(108), 103-105.
- Kim, H., & Park, Y. (2012). The complete mitochondrial genome of the Korean red-backed vole, *Myodes regulus* (Rodentia, Murinae) from Korea. *Mitochondrial DNA*, 23(2), 148-150.
- Kimura, M. (1983). *The neutral theory of molecular evolution*. Cambridge [Cambridgeshire: Cambridge University Press.
- Kosakovsky Pond, S., & Frost, S. (2005a). A Genetic Algorithm Approach to Detecting Lineage-Specific Variation in Selection Pressure. *Mol. Biol. Evol.*, 22(3), 478-485.
- Kosakovsky Pond, S., & Frost, S. (2005b). Not So Different After All: a Comparison of Methods for Detecting Amino Acid Sites Under Selection. *Mol. Biol. Evol.*, 22(3), 1208-1222.

- Kotlík, P., Deffontaine, V., Mascheretti, S., Zima, J., Michaux, J., & Searle, J. (2006). A northern glacial refugium for bank voles (*Clethrionomys glareolus*). *Proc. Natl. Acad. Sci. U. S. A.*, *103*(40), 14860-14864.
- Kotlík, P., Marková, S., Vojtek, L., Stratil, A., Šlechta, V., Hyršl, P., & Searle, J. (2014). Adaptive phylogeography: Functional divergence between haemoglobins derived from different glacial refugia in the bank vole. *Proc. R. Soc. B*, *281*(1786), 1-9.
- Kubli, D., & Gustafsson, A. (2012). Mitochondria and Mitophagy: The Yin and Yang of Cell Death Control. *Circ. Res.*, *111*(9), 1208-1221.
- Kvist, L., Martens, J., Nazarenko, A., & Orell, M. (2003). Paternal Leakage of Mitochondrial DNA in the Great Tit (*Parus major*). *Mol. Biol. Evol.*, *20*(2), 243-247.
- Lanfear, R., Calcott, B., Ho, S., & Guindon, S. (2012). PartitionFinder: Combined Selection of Partitioning Schemes and Substitution Models for Phylogenetic Analyses. *Mol. Biol. Evol.*, *29*(6), 1695-1701.
- Lopez, J., Yuhki, N., Masuda, R., Modi, W., & O'Brien, S. (1994). Numt, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat. *J Mol Evol.*, *39*(2), 174-90.
- Lynch, M. (1996). Mutation accumulation in transfer RNAs: Molecular evidence for Muller's ratchet in mitochondrial genomes. *Mol. Biol. Evol.*, *13*(1), 209-220.
- Macalpine, D., Kolesar, J., Okamoto, K., Butow, R., & Perlman, P. (2001). Replication and preferential inheritance of hypersuppressive petite mitochondrial DNA. *EMBO J.*, *20*(7), 1807-1817.
- Marchington, D., Hartshorne, G., Barlow, D., & Poulton, J. (1997). Homopolymeric Tract Heteroplasmy in mtDNA from Tissues and Single Oocytes: Support for a Genetic Bottleneck. *Am. J. Hum. Genet.*, *60*, 408-416.
- Marková, S., Dufresne, F., Manca, M., Kotlík, P., & Fontaneto, D. (2013). Mitochondrial Capture Misleads about Ecological Speciation in the *Daphnia pulex* Complex. *PLoS ONE*, E69497-E69497.
- Marková, S., Searle, J., & Kotlík, P. (2014). Relaxed functional constraints on triplicate α -globin gene in the bank vole suggest different evolutionary history from other rodents. *Heredity*, *113*, 64-73.

- McClellan, D., Palfreyman, E., Smith, M., Moss, J., Christensen, R., & Sailsbery, J. (n.d.). Physicochemical Evolution And Molecular Adaptation Of The Cetacean And Artiodactyl Cytochrome B Proteins. *Mol. Biol. Evol.*, 22(3), 437-455.
- McGowen, M., Grossman, L., & Wildman, D. (2012). Dolphin genome provides evidence for adaptive evolution of nervous system genes and a molecular rate slowdown. *Proc. R. Soc. B*, 279(1743), 3643-3651.
- Mitchell, P. (1961). Coupling of Phosphorylation to Electron and Hydrogen Transfer by a Chemi-Osmotic type of Mechanism. *Nature*, 191(4784), 144-148.
- Moses, M. (1961). Spermiogenesis in the crayfish (*Porcamabarus clarkii*): II. Description of Stages. *J. Cell Biol.*, 10(3), 301-333.
- Nachman, M., Brown, W., Stoneking, M., & Aquadro, C. (1996). Nonneutral Mitochondrial DNA Variation in Humans and Chimpanzees. *Genetics.*, 142(3), 953–963.
- Nass, S., & Nass, M. (1963). Intramitochondrial fibers with DNA characteristics. *J Cell Biol.*, 19(3), 593-629.
- Nečas, O., Hejtmánek, M., Janisch, R., Červinka, M., Lenhart, K., & Kolář, Z. (2000). Membránový princip funkční organizace buňky - Mitochondrie a chloroplasty. In *Obecná biologie* (3rd ed., p. 205). Jinočany: H&H Vyšehradská, s. r. o.
- Ohta, T. (2002). Near-neutrality in evolution of genes and gene regulation. *Proc. Natl. Acad. Sci. U. S. A.*, 99(25), 16134-16137.
- Parmakelis, A., Kotsakiozi, P., Rand, D., & Bourtzis, K. (2013). Animal Mitochondria, Positive Selection and Cyto-Nuclear Coevolution: Insights from Pulmonates. *PLoS ONE*, 8(4), E61970-E61970.
- Rato, C., Carranza, S., & Harris, D. (2011). When selection deceives phylogeographic interpretation: The case of the Mediterranean house gecko, *Hemidactylus turcicus* (Linnaeus, 1758). *Mol. Phylogenet. Evol.*, 58(2), 365-373.
- Rea, S. (2005). Metabolism in the *Caenorhabditis elegans* Mit mutants. *Exp. Gerontol.*, 40(11), 841-849.
- Ronquist, F., Huelsenbeck, J., & Teslenko, M. (2011). Draft MrBayes version 3.2 Manual: Tutorials and Model Summaries. 21-21. Retrieved July 22, 2014, from

http://mr bayes.sourceforge.net/mb3.2_manual.pdf

- Schnare, M., & Gray, M. (1982). 3'-Terminal sequence of wheat mitochondrial 18S ribosomal RNA: Further evidence of a eubacterial evolutionary origin. *Nucleic Acids Res.*, 10(13), 3921-3932.
- Searle, J., Kotlík, P., Rambau, R., Marková, S., Herman, J., & Mcdevitt, A. (2009). The Celtic fringe of Britain: Insights from small mammal phylogeography. *Proc. R. Soc. B*, 276(1677), 4287-4294.
- Shaw, K. (2002). Conflict between nuclear and mitochondrial DNA phylogenies of a recent species radiation: What mtDNA reveals and conceals about modes of speciation in Hawaiian crickets. *Proc. Natl. Acad. Sci. U. S. A.*, 99(25), 16122-16127.
- Silvestro, D., & Michalak, I. (2012). RaxmlGUI: a graphical front-end for RAxML. *Org. Divers. Evol.*, 12(4), 335-337.
- Soares, P., Pereira, L., Samuels, D., Macaulay, V., Li, B., Radivojac, P., ... Caramelli, D. (2013). Evaluating Purifying Selection in the Mitochondrial DNA of Various Mammalian Species. *PLoS ONE*, 8(3), E58993-E58993.
- Sommer, R., & Nadachowski, A. (2006). Glacial refugia of mammals in Europe: Evidence from fossil records. *Mamm. Rev.*, 36(4), 251-265.
- Song, S., Pursell, Z., Copeland, W., Longley, M., Kunkel, T., & Mathew, C. (2005). DNA precursor asymmetries in mammalian tissue mitochondria and possible contribution to mutagenesis through reduced replication fidelity. *Proc. Natl. Acad. Sci. U. S. A.*, 102(14), 4990-4995.
- Sorenson, M., & Quinn, T. (1998). Numts: a Challenge for Avian Systematics and Population Biology. *The Auk*, 115(1), 214-221.
- Stamatakis, A., Ludwig, T., & Meier, H. (2005). RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. *Bioinformatics*, 21(4), 456-463.
- Stamatakis, A., & Alachiotis, N. (2010). Time and memory efficient likelihood-based tree searches on phylogenomic alignments with missing data. *Bioinformatics*, 26(12), I132-I139.
- Stojanovski, D., Johnston, A., Streimann, I., Hoogenraad, N., & Ryan, M. (2003).

Import of nuclear-encoded proteins into mitochondria. *Exp. Physiol.*, 88(1), 57-64.

- Sutovsky, P., Moreno, R., Ramalho-Santos, J., Dominko, T., Simerly, C., & Schatten, G. (1999). Ubiquitin tag for sperm mitochondria. *Nature*, 402, 371–372.
- Taberlet, P., Fumagalli, L., Wust-Saucy, A., & Cosson, J. (1998). Comparative phylogeography and postglacial colonization routes in Europe. *Mol. Ecol.*, 7(4), 453-464.
- Tait, S., & Green, D. (2012). Mitochondria and cell signalling. *J. Cell Sci.*, 125(4), 807-815.
- Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Mol. Biol. Evol.*, 30(12), 2725-2729.
- Tatarenkov, A., & Avise, J. (2007). Rapid concerted evolution in animal mitochondrial DNA. *Proc. R. Soc. B*, 274(1619), 1795-1798.
- Tegelström, H. (1987). Transfer of mitochondrial DNA from the northern red-backed vole (*Clethrionomys rutilus*) to the bank vole (*C. glareolus*). *J. Mol. Evol.*, 24(3), 218-227.
- Temperley, R., Wydro, M., Lightowers, R., & Chrzanowska-Lightowers, Z. (2010). Human mitochondrial mRNAs—like members of all families, similar but different. *BBA - Bioenergetics*, 1797(6-7), 1081-1085.
- Triant, D., & Dewoody, J. (2007). The Occurrence, Detection, and Avoidance of Mitochondrial DNA Translocations in Mammalian Systematics and Phylogeography. *J. Mammal.*, 88(4), 908-920.
- Turelli, M., Hoffmann, A., & McKechnie, S. (1999). Dynamics of cytoplasmic incompatibility and mtDNA variation in natural *Drosophila simulans* populations. *Genetics*, 132(3), 713-723.
- Ursprung, H., & Schabtach, E. (1965). Fertilization in tunicates: Loss of the paternal mitochondrion prior to sperm entry. *J. Exp. Zool.*, 159(3), 379-383.
- Vawter, L., & Brown, W. (1986). Nuclear And Mitochondrial DNA Comparisons Reveal Extreme Rate Variation In The Molecular Clock. *Science*, 234(4773), 194-196.
- Warburg, O., & Mayerhof, O. (1914). (). *Pfluegers Archiv Fuer Die Gesamte*

Physiologie Des Menschen Und Der Tiere., 148, 295-310.

- Woolley, S., Johnson, J., Smith, M., Crandall, K., & McClellan, D. (2003). TreeSAAP: Selection on Amino Acid Properties using phylogenetic trees. *Bioinformatics*, 19(5), 671-672.
- Wójcik, J., Kawałko, A., Marková, S., Searle, J., & Kotlík, P. (2010). Phylogeographic signatures of northward post-glacial colonization from high-latitude refugia: a case study of bank voles using museum specimens. *J. Zool.*, 281, 249-262.
- Yang, D., Oyaizu, Y., Oyaizu, H., Olsen, G., & Woese, C. (1985). Mitochondrial origins. *Proc. Natl. Acad. Sci. USA*, 82, 4443-4447.
- Yasukawa, T., Reyes, A., Cluett, T., Yang, M., Bowmaker, M., Jacobs, H., & Holt, I. (2006). Replication of vertebrate mitochondrial DNA entails transient ribonucleotide incorporation throughout the lagging strand. *EMBO J.*, 25(22), 5358-5371.
- Yu, X., & Gabriel, A. (1999). Patching Broken Chromosomes with Extranuclear Cellular DN
- Zwickl, D. (2006). *Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion*. Austin, Tex.: University of Texas Libraries.